

Continual Process Improvement based on Reference Models and Process Mining

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)
im Fach Wirtschaftswissenschaft

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von

Dipl.-Betriebswirtin (FH) Kerstin Gerke
geboren am 05.03.1970 in Dortmund

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Oliver Günther, Ph.D.

Gutachter:

1. Prof. Oliver Günther, Ph.D.
2. Prof. Dr. Myra Spiliopoulou

eingereicht am: 07.02.2011

Tag der mündlichen Prüfung: 19.05.2011

Preface

Information systems are becoming more and more intertwined with the operational processes they support. As a result, multitudes of events are recorded by today's information systems. Nevertheless, organizations have problems extracting value from these data. The goal of process mining is to use event data to extract process-related information, for example, to automatically discover a process model by observing events recorded by some enterprise system. Process mining is a relative young research discipline that sits between machine learning and data mining on the one hand and process modeling and analysis on the other hand.

The idea of process mining is to discover, monitor, and improve real processes (i. e., not assumed processes) by extracting knowledge from event logs readily available in today's systems. In my view, this is by far the most exciting and innovative development in BPM research that has its roots in office automation and workflow management. Therefore it is a real pleasure to write the preface for this dissertation. Kerstin Gerke did not focus on new process algorithms, but focused on the application of these techniques and the embedding of process mining in the "bigger picture" of process management and improvement.

In her dissertation, Kerstin Gerke describes several case studies. This provides an empirical analysis of process mining. Moreover, she shows the opportunities and challenges of using new data carriers such as RFID. She also positions process mining in the context of ITIL and the whole BPM life cycle and presents an integrated business process control framework. In the last part of her thesis she focuses on compliance issues thereby linking reference models to event logs. This is a hot topic in process mining research and it is clear that this will trigger important innovations in corporate governance and auditing.

I am sure that you will enjoy reading this thesis and distill actionable knowledge from it. I hope that it will trigger the reader to apply process mining technology and thus improve processes. To conclude, I would like to congratulate Kerstin and her supervisors with the completion of this nice thesis. It must have been difficult to work on a PhD thesis while having an industry job and being a young mother. Fortunately, the satisfaction of completion typically strongly correlates with the efforts put into it.

Prof.dr.ir. Wil van der Aalst
Eindhoven University of Technology

Abstract

The dissertation at hand takes as its subject business processes. Naturally they are subject to continual improvement and are a major asset of any given organization. An optimally-designed process, having once proven itself, must be flexible, as new developments demand swift adaptations.

However, many organizations do not adequately describe or utterly fail to describe these processes, though doing so is a prerequisite for their improvement. Very often the process model created during an information system's implementation either is not used in the first place or is not maintained, resulting in an obvious lack of correspondence between the model and operational reality. Process mining techniques prevent this. They extract the process knowledge inherent in an information system and visualize it in the form of process models. Indeed, continual process improvement depends greatly on this modeling approach, and reference models, such as ITIL and COBIT, are entirely suitable and powerful means for dealing with the efficient design and control of processes.

Process improvement typically consists of a number of analysis, design, implementation, execution, monitoring, and evaluation activities. This dissertation proposes a methodology that supports and facilitates them. A procedural model is used that continually controls the correspondence of the real processes with both business requirements and IT systems, thereby identifying, or helping to identify, potential improvements.

An empirical analysis both revealed the challenges and the potential benefits of these processes mining techniques' successful use and spurred the development of new analysis in this domain. This in turn led to the detailed consideration of specific aspects of the data preparation for process mining algorithms. Here the focus is on the provision of enterprise data and RFID events.

This dissertation as well examines the importance of analyzing the execution of reference processes to ensure compliance with modified or entirely new business processes. Moreover, a contribution to the development of a compliance application is made.

The methodology involved a number of cases' practical trials; the results demonstrate its power and universality. This new approach ushers in an enhanced continual inter-departmental and inter-organizational improvement process.

Keywords: Compliance, Business processes, Reference models, Process mining

Zusammenfassung

Geschäftsprozesse, mit deren kontinuierlicher Verbesserung sich die vorliegende Arbeit befasst, stellen ein wichtiges Gut eines Unternehmens dar. Für den Unternehmenserfolg sind nicht einmalig optimal gestaltete Prozesse entscheidend, sondern die Fähigkeit, schnell auf neue Entwicklungen reagieren und die betroffenen Prozesse flexibel anpassen zu können. In vielen Unternehmen ist eine aktuelle Beschreibung ihrer Prozesse als wesentliche Voraussetzung für die Prozessverbesserung jedoch nicht oder nur unzureichend gegeben. Nicht selten wird ein mit einer Systemeinführung erstelltes Prozessmodell nicht weiterverwendet und gepflegt, so dass es nach kurzer Zeit von der betrieblichen Realität abweicht. Diese fehlende Übereinstimmung kann durch die Nutzung von Prozess-Mining-Technologien verhindert werden, indem das in den Informationssystemen implizit vorhandene Prozesswissen automatisiert extrahiert und in Form von Prozessmodellen abgebildet wird. Eine zentrale Annahme dieser Arbeit ist, dass dieser Modellierungsansatz eine hohe Relevanz für die kontinuierliche Prozessverbesserung aufweist. Eine weitere zentrale Annahme ist, dass Referenzmodelle, wie z. B. ITIL und COBIT, ein wichtiges Element für die effiziente Gestaltung und Steuerung von Prozessen bilden.

Die Prozessverbesserung durchläuft in der Regel mehrere Analyse-, Design-, Implementierungs-, Ausführungs-, Monitoring-, und Evaluierungsschritte. Die Arbeit stellt eine Methodik vor, die die Identifizierung und Lösung der auftretenden Aufgaben unterstützt und erleichtert. Dabei wird ein Vorgehensmodell genutzt, das die Übereinstimmung der tatsächlich “gelebten” Prozesse mit den Geschäftsanforderungen und den IT-Systemen überwacht und bei der Identifizierung möglicher Verbesserungspotenziale unterstützt.

Eine empirische Untersuchung zeigt die Herausforderungen und die Potenziale für den erfolgreichen Einsatz von Process-Mining-Techniken auf und motiviert die Entwicklung neuer Analysetechniken in diesem Bereich. Auf der Basis der Resultate dieser Untersuchung wurden spezielle Aspekte der Datenaufbereitung für Process-Mining-Algorithmen detailliert betrachtet. Der Fokus liegt dabei auf der Bereitstellung von Enterprise- und RFID-Daten.

Weiterhin beleuchtet die Arbeit die Wichtigkeit, die Referenzprozessausführung zu überprüfen, um deren Einhaltung in Bezug auf neue oder geänderte Prozesse zu sichern. Zudem stellt sie den eigenen Beitrag zur Entwicklung einer Compliance-Anwendung vor.

Die Methodik wurde anhand einer Reihe von Praxisbeispielen erprobt. Die Ergebnisse unterstreichen ihre generelle unternehmensübergreifende Anwendbarkeit für die effiziente kontinuierliche Prozessverbesserung.

Schlagworte: Compliance, Geschäftsprozesse, Referenzmodelle, Process-Mining

*To Marcus without whose love, encouragement, and support, this dissertation and very
much more would never have come to be.*

Contents

1. Introduction	1
1.1. Motivation	1
1.2. Problem Statement	4
1.3. Principal Contributions	6
1.4. Outline	8
2. Process Organizations	11
2.1. Business Process Management	11
2.1.1. Background	11
2.1.2. Definitions of Business Process Management	14
2.2. Business Process Management Concepts	15
2.2.1. Business Process Engineering	15
2.2.2. Business Process Reengineering	18
2.2.3. Continuous Process Improvement	19
2.3. Process Modeling	19
2.3.1. Petri Net	20
2.3.2. Event-Driven Process Chain	22
2.3.3. Business Process Management Notation	23
2.4. Process Engineering and Reengineering with Reference Models	23
2.4.1. Definition of Reference Models	24
2.4.2. Domains	25
2.4.3. Reference Models	25
2.5. Process Reverse Engineering with Process Mining	29
2.5.1. Definition of Process Mining	29
2.5.2. Perspectives	30
2.5.3. Representatives	31
2.5.4. Process Mining Facilities	32
2.6. Technology Support for Process Organizations	33
2.6.1. Background and Definitions	33
2.6.2. Value of Information Technology	34
2.6.3. IT-Business Alignment	35
2.6.4. Compliance and Risk	38
2.6.5. IT Governance	39
2.6.6. RFID Technology	41
2.6.7. EPCglobal	42

3. Exploiting New Data Sources for Process Mining	47
3.1. Related Work	47
3.2. Data Preparation	49
3.2.1. Meta Model for Process Mining Data	50
3.2.2. Requirements	51
3.3. Mining Supply Chain Processes	52
3.3.1. Challenges of Mining EPCIS Events	53
3.3.2. Preparation of EPCIS Events for Process Mining	55
3.4. Mining Enterprise Processes	63
3.4.1. Challenges of Mining Enterprise Processes	63
3.4.2. Preparation of Enterprise Data for Process Mining	67
3.5. Conclusion	79
3.5.1. Process Discovery in Enterprise Systems	79
3.5.2. Process Discovery in RFID Environments	80
4. Process Discovery with Process Mining	83
4.1. Case Studies	83
4.1.1. Car Production and Delivery	84
4.1.2. Pharmaceutical Drug Delivery	89
4.1.3. Complaint Handling	91
4.1.4. Service Operation	92
4.2. Process Discovery	95
4.2.1. Car Production and Delivery	96
4.2.2. Pharmaceutical Drug Delivery	108
4.2.3. Complaint Handling	114
4.2.4. Service Operation	119
4.3. Conclusion	124
5. Business Process Control Framework	131
5.1. Related Work	131
5.1.1. Event Log Analysis	131
5.1.2. Process Controlling	132
5.1.3. Data Warehouse Concepts	132
5.2. Concept of Continual Process Improvement	132
5.2.1. Procedure Model for Service Operation Based on ITIL	133
5.2.2. Procedure Model for Continual Service Improvement Based on ITIL	133
5.2.3. Procedure Model for Service Operation and Continual Service Im-	
provement Based on Process Mining	134
5.3. Definition of a Business Process Control Framework	135
5.3.1. Need for a New Control Framework	135
5.3.2. Relevant Entities	142
5.3.3. Procedure Model for IT Service Operation Based on ITIL and	
Process Mining	143
5.3.4. Compatibility with Reference Models	146

5.4.	Continual Process Improvement in Practice	148
5.4.1.	Complaint Handling	148
5.4.2.	Service Operation	151
5.5.	Conclusion	160
6.	Measuring Process Compliance with Reference Models	163
6.1.	Related Work	163
6.1.1.	Equivalence Notions	164
6.1.2.	Similarity Notions	165
6.1.3.	String Matching	166
6.2.	Requirements of Measuring Compliance	167
6.2.1.	Compliance Maturity and Degree	168
6.2.2.	Granularity of Models	168
6.2.3.	Customization of the Reference Model	169
6.2.4.	Iteration of Process Activities	169
6.3.	Measuring Process Compliance with Quality Indicators	170
6.3.1.	Log-Based Analysis	171
6.3.2.	Inter-Model Analysis	173
6.4.	Development of a Sequence-Based Compliance Analysis	175
6.4.1.	Theoretical Foundations	175
6.4.2.	Measuring Compliance	177
6.4.3.	Feasibility Study	179
6.4.4.	Running Example	181
6.5.	Measuring Compliance in Practice	185
6.5.1.	Complaint Handling	185
6.5.2.	Service Operation	189
6.5.3.	Pharmaceutical Drug Delivery	190
6.6.	Conclusion	193
7.	Conclusion	195
7.1.	Results	195
7.2.	Lessons Learned	196
7.3.	Significance of Findings	197
7.4.	Future Issues and Directions	199
	Appendices	203
A.	Questionnaire	203
A.1.	Importance and Performance of IT and IT Management	204
A.2.	Alignment of IT Management and Business Process Management	206
A.3.	Effectiveness of Reference Models	208
A.4.	Effectiveness of Process Mining	210

Contents

B. Analysis	213
B.1. Importance and Performance of IT and IT Management	213
B.2. Alignment of IT Management and Business Process Management	215
B.3. Effectiveness of Reference Models	217
B.4. Effectiveness of Process Mining	219
C. Process Models	223
D. Process Mining Algorithms	231
E. SCOR D2 Process Strategy	237
Bibliography	239
Author's Publications	259
List of Figures	261
List of Tables	265
Acronyms	267
Acknowledgment	271

1. Introduction

This dissertation has as its subject the continual improvement of processes using process mining and reference models. Hereafter, Section 1.1 describes the paper's inspiration. After a discussion of general problems in Section 1.2, Section 1.3 describes the principal research contributions by discussing the major findings. Finally, Section 1.4 provides an outlook of the paper's structure.

1.1. Motivation

From the early 1990's business processes have established their relevance as a major asset of an organization (Leymann and Altenhuber, 1994). They are the means by which the organization accomplishes its competencies in generating value for the customer. Efficient business processes are therefore a vital factor in the financial success. Information technology (IT) plays a key role¹ (IT Governance Institute, 2008a, p. 15 ff.; IT Governance Institute, 2009, p. 19 ff.) in achieving process efficiency (Herbst and Karagiannis, 1998; Porter and Millar, 1985, p. 151), and organizations therefore are becoming increasingly dependent on IT to run profitable businesses (Office of Government Commerce, 2007).

Information technology service management (ITSM) has met the major challenge of supporting business processes by producing and maintaining high-quality IT services. IT service is offered to internal customers or external customers, or to both of them, and is provided to users by either an internal IT department or an external IT service provider. Both are here referred to as IT service provider. ITSM further faces the challenge of successfully aligning these services with business requirements. Empirical studies show that successful alignment can lead to a higher value contribution (Chan et al., 1997; Tallon et al., 2000). Chan et al. (1997) observed that a high degree of alignment raises both the effectiveness and the efficiency of the organization.

The integrated view of business and IT has strengthened the practice of service management as well as imposing greater challenges on IT (Office of Government Commerce, 2007). In today's highly competitive commercial world, IT service providers are judged on their ability to deliver high-quality IT services, this in a business environment that often increases the frequency, complexity, and the extent of changes apace. Thus, business requirements compel the utmost flexibility with respect to IT services and customer demands.

¹ Incidentally, the value of IT has for years provoked a great deal of debate and controversy – this will be fully addressed in Section 2.6.2.

1. Introduction

Because of these pressures, quality enhancement has become mainstream thinking for IT service providers. It is important that this commitment to quality is targeted both to the production processes of IT services and to those of the ITSM. The reference model Control Objectives for Information and related Technology (COBIT) evinces this. According to COBIT, a prerequisite for a high maturity² level is the automation of both that of the operational processes and that of the process management activities (IT Governance Institute, 2007). An examination by the London School of Economics reaches this very conclusion, pointing out that the management of technology is decisive to guarantee its efficient and effective use (van Reenen and Sadun, 2005).

However, scrutiny of the instruments for the management of IT reveals that they are far less established than other management areas (Johannsen and Goeken, 2007, p. 2). Increasingly, reference models, such as ITIL and COBIT, are found in this environment. The growing interest in reference modeling for ITSM is due to its substantial contribution to designing, operating, and controlling ITSM processes efficiently. In general, reference models offer the chance to improve processes and management structures because ITSM methods and procedures follow proven standards that save both time and effort. There are several incentives for adopting them. First, they optimize the design because they have been developed, usually incorporating the insights of experts, over a long period. Second, they significantly speed up the design of process models by providing reusable and high quality content. Third, they ease compliance³ with industry regulations. Fourth, they are an essential means of creating a link between business needs and IT implementation (van der Aalst et al., 2006).

The absence of a method and the use of an outdated one explain why IT service providers often fall back on business process management (BPM) methods and tools – they support the analysis, simulation, enactment, and continual change of business processes, and commensurately contribute to business improvement.

Having outlined the interaction of business and IT, attention is now turned to business improvement of IT-enabled processes⁴. As earlier noted, business advantages can often be gained through process efficiencies, specifically through both ITSM process and business process efficiencies. That is, one way to optimize an extant business process is to optimize its IT-support.

Note initially that process improvement requires a thorough analysis of present business processes (henceforth “as-is analysis”). Business process modeling techniques are frequently used formally to describe business processes. Customarily, the as-is analysis of business processes is carried out by interviewing persons, called process experts, who are actively involved in the business processes’ execution. It may well be that the description of the business process is strongly based on the opinions which may vary substantially from one person to the next due to conflicting interests of these process

² Note that COBIT refers to the quality of a process as to process maturity.

³ Compliance is defined as the degree to which the process is implemented as described and its results are in accordance with the laws, regulations, and contractual arrangements to which the business process is subject, that is, externally imposed business criteria as well as internal policies and standards (IT Governance Institute, 2007).

⁴ IT-enabled processes are processes that are supported by IT.

experts – or, to frame it differently, on *knowledge*, which may be highly diffused within an organization, and which may be conveyed with ambiguous utterances even when describing the same processes. Modeling errors are commonly detected only after a process model is implemented (Herbst and Karagiannis, 1998), further confounding matters. As the acquisition of the initial process model is a complex and therefore lengthy and generally expensive endeavor, the customary as-is analysis of BPM is quite limited for the acquisition of process models - hence many organizations' process descriptions are either inadequate or entirely unavailable.

It is safe to assume that numerous companies resist change (zur Muehlen, 2004; Mansir and Schacht, 1989, Chapter 1, p. 3). As the result of this resistance, a process model may no longer correspond to the operational reality. Morgan and Schiemann (1999) have already stressed that outdated metrics or those misaligned with organizational objectives can diminish benefits. Also, implementation within the information systems (ISs) may vary, despite employees' efforts to the contrary, from the documented to-be processes, most starkly as to reference models.

Once the processes are formally described, the second step within process improvement can commence: The investigation of weak points and improvement potentials. The as-is processes are compared with the to-be processes, revealing those elements that prevent objectives to be reached efficiently. Naturally, knowledge of discrepancies can lead to suggestions for improving the as-is processes, as can models showing the degree of process maturity. Based on these findings, to-be models of the improved business processes are made and implemented (Allweyer and Scheer, 1995).

After the processes have been designed or redesigned according to the reference model, it is necessary continuously to monitor process execution – it is analyzed after the fact. In order to identify possible quality problems, organizations commonly measure the efficiency and effectiveness of their ITSM processes with key indicators. Target value compliance anomalies suggest that a process goal is in jeopardy (Österle, 1995, p. 18).

Once having clearly defined processes in place, organizations ought to advance the automation of their ITSM processes and make use of both emerging technologies and reference models to reach the next maturity level.

Process mining, for example, facilitates the analysis of processes by extracting a process model from event logs kept during the execution of IT-enabled business processes. Process mining thereby potentially bridges the gap between the description of processes in model-based representations and their actual performance in information systems. By increasing the transparency of the as-is situation, process mining promises to alleviate a plethora of business problems. Modeling becomes independent from the process experts' conceptions. A business availing itself of the potential of automating the process analysis might vastly improve cost and time factors for BPM, challenging the traditional process modeling stage. Therefore, the question arises: How can process mining be seamlessly integrated into ITSM processes to improve continuously ITSM processes, and with that contribute to business process improvement? There is a clear need for an in-depth analysis of process mining benefits and an evaluation of process mining use cases. In addition to studying the benefits, the potential liabilities, which may prevent the universal use of process mining one day, must be appraised.

1. Introduction

Obviously, reference models might be useful in improving processes but the question arises to what extent these best practices can be adopted and implemented in a specific business context – a question implicating compliance with reference models. Process mining has shown considerable potential for the comparison of any two processes. Consequently, the next step is ascertaining whether process mining can be exploited to assess the compliance of processes with reference models – a study of the specifics of reference models is needed.

Considering that currently substantial effort is being put into continuously improving the process quality, clearly it is equally important to develop further the continuous improvement process itself.

1.2. Problem Statement

This section introduces major problems and discusses controversies that arise from the issues identified. In general, action is *slowly* being taken to implement IT management activities (IT Governance Institute, 2008a, p. 34). Take IT reference models; they have been established in practice just recently. However, not every development has been discouraging. A large increase in the adoption and use of COBIT, for instance, is evident (from ten percent in 2006 to about thirty percent in 2009) (IT Governance Institute, 2008a, p. 36)); the great attention these models have garnered and their broad acceptance suggest that the IT had worked haphazardly rather than systematically so far. IT management methods are seldom applied in practice, an unsurprising reality given the paucity of their discussion in theory, that is, in computer science circles. The methods of IT robustly address application and system development as well as the operative business, but perhaps faintly address management aspects (Johannsen and Goeken, 2007, p. 2).

Supporting BPM, specifically improving the IT efficiency with process mining and reference models in light of process improvement, addresses at least six current challenges organizations face.

First, IT management needs guidance on implementing efficient IT processes, which support the business requirements. Granted, reference models promise instrumental support in creating value for customers through better design, transition, and operation of processes. Nevertheless, reference models need to be adapted to organizational requirements (Reijers, 2005). This tailoring is necessary because reference models are fairly generic and they are often only partially applied or they are used in tandem with other best practices (Kütz, 2009). This is exemplified by the combination of ITIL with COBIT. Guidance particularly has to be given as to what extent the reference models are adapted and implemented in a specific business context.

Second, IT management needs potent support for objective decision making as to acquiring and adapting process models as a reflection of the as-is situation. Traditional process modeling, however, remains quite limited because it is time-consuming, fault-prone, and generally expensive. It is exactly for these reasons that one of the most time-consuming activities in process improvement is the acquisition of the initial process

model (Herbst and Karagiannis, 1998). Still, an accurate account of what is happening in the organization remains a prerequisite for process improvement.

Third, IT management has few methods for measuring the discrepancy between how business activities are actually performed and how they ought to be carried out in ISs. This problem encompasses the accuracy of processes and their model-based representations, IT implementations, the alignment of IT and business, and the effectiveness of the applications in use. Although the management information system (MIS) has evolved much since the 1950's, very few versions of it offer methods that are capable of analyzing the performance of IT-enabled processes, an analysis that when done properly takes the business perspective into account (Herbst and Karagiannis, 1998; zur Muehlen, 2004, p. 5). It is important to stress that management information systems typically provide information about key performance indicators (KPIs) to measure the performance of their IT processes but not about the underlying business processes (Dumas et al., 2005, p. 237). This does not imply that the key performance indicators are no longer needed. They are certainly important to substantiate managerial decisions; but since the processes themselves represent important information resources for identifying weak points and potential improvements (Mansir and Schacht, 1989, p. 3-6), it is obvious that methods, techniques, and tools to analyze, manage, and support business processes are also understandably in demand (Leymann and Altenhuber, 1994; zur Muehlen, 2004). "The lack of good white-collar measurements is", as Harrington (1991) phrased it, "a major obstacle to improve business processes. [...] if you cannot measure it, you cannot control it. And if you cannot control it, you cannot manage it."

Fourth, IT management is often unaware of the extent to which organic growth, acquisitions, or changing business needs have imperceptibly resulted in numerous ways of executing similar activities (Object Management Group, 2008). Variations in process execution bring about a gap between the implementation within the information systems and the documented processes. More worryingly, it is to assume that the process variants unnecessarily increase the complexity of the IS, which in turn results in opacity of cause-effect relationships that describe the effects of changes on the process level in information systems (zur Muehlen, 2004, p. 5), thus entrenching resistance toward change.

Fifth, IT management needs processes that are based on an IT infrastructure in line with business requirements. In practice, however, using IT in a way that really contributes effectively to the achievements of business is frequently problematic. Although progress in aligning business processes with IT is discernible (IT Governance Institute, 2009, pp. 10–11), there is still substantial room for improvement (IT Governance Institute, 2008a, p. 26). A stark example: In the vast majority of organizations IT architectures determine the business processes, not vice-versa (Acrys Consulting, 2005). Troubling too is the frequent absence of methodical support for this task (Avison et al., 2004). The same applies to IT governance, which identifies alignment as one of the five focus areas along with risk management, resource management, value delivery, and performance measurement (IT Governance Institute, 2008a, p. 19 ff.).

Last and perhaps most distressing the degree of automation in the active handling and development of ITSM processes remains dismal. Key sources of problems are missing

1. Introduction

or unexploited tools between the various perspectives and the various stages in the life cycles of processes. Typical examples are incompatible representations of the managerial perspective versus the working level perspective and the gap between normative modeling for compliance purposes and the actual execution of a process.

Even if process mining contributes to business process automation – most process mining algorithms do perform well on single-system event logs that explicitly refer to a process instance, also known as a case – in many operational environments such case identifiers are not directly recorded for events. The missing identifiers limit a comprehensive use of process mining. In supply chain processes there are further challenges, since different identification numbers, vertical integration, and numerous aggregation steps prevent individual work steps from becoming traceable as a case. As a result, there are few instances of the use of process mining in supply chains.

1.3. Principal Contributions

This dissertation covers the complete spectrum from problem domain analysis to the discussion of an appropriate methodology for continually improving processes to the application of this methodology. A number of implementations are successfully carried out as proof of the concept. Their results demonstrate the viability of the methodology developed. The following is a categorization and a summary of the major contributions.

1. *Conceptual contributions* empirically illustrating the situation in the business process reengineering (BPR) domain and resulting in the *continual process and service improvement (CPSI)* methodology. The concept is mainly inspired by five domain-spanning case studies, which were conducted at diverse organizations. The case study research method has been chosen as a coherent research strategy for clearly understanding real-life challenges to process management. Along with these case studies, the methodology that actively supports the analysis, the monitoring, and the control process for continually improving processes⁵ is developed and discussed.

The use of process mining as well as that of reference models is a specific means to enable BPR. It is shown that both process mining and the use of reference models are valid and powerful techniques. This is why the proposed methodology emphasizes the interconnection of reference models and process mining. This has been chosen not only to ensure that the business processes and the ITSM processes function exactly as specified but to identify and correct service-specific weaknesses of the process implementation. It further integrates ITSM processes, staff, and resources into the CPSI methodology and efficiently deals with the multitude of inevitable management tasks, such as process modeling, documenting, and analyzing.

2. *Design contributions* developing solutions for selected problems associated with the CPSI explicitly target solutions that are requested but as yet unmet. The

⁵ The term *process* is here understood in its broadest sense to cover all phenomena from which a traceable instance can be identified: activities, states, relations, or events, whether physical or not.

focus is on avoiding “re-inventing” concepts and methods already extant, thereby benefiting from adaptation, combination, and integration.

Despite potential benefits, reference models and process mining are not yet widely adopted in industry, particularly not in tandem. The impediments to deep penetration are analyzed in-depth based on experiences from the case studies and surveys. They tend to disclose that uncertainty about the benefits, missing methods, and even at times obliviousness to the process mining itself conspire to lead to low diffusion in industry. It is shown here how to overcome these liabilities. The most striking result of this dissertation may be a procedure model for the CPSI, one which manifests deviations from reference models through continually monitoring the behavior of the process execution and verifying it against specifications. It is shown that the deviations follow patterns that indicate weak points either in the process per se or in the process implementation. Thus, changes in the process execution from growth, acquisitions, or changing business needs become perceptible. It is further demonstrated that the integrated use of process mining with reference models is particularly suitable to improve the maturity level of ITSM processes, that it contributes to IT governance, and that it offers considerable potential for automation.

Given the fact that companies may fail to take into account factors and characteristics related to compliance of processes with reference models, compliance is identified as a substantial functional requirement for verifying processes within methodology. It is demonstrated that process models can have different structures but one process can still be compliant with the other. Therefore, the characteristics of compliance are investigated in more detail. As an important and perhaps ground breaking result of this part of the paper new measurements are provided expressing compliance with reference models. The strength of the quality indicators lies in explaining how reference models are adapted and implemented in a specific business context. The value of providing such information also has markedly positive effects on IT governance, the measurements showing their relevance commensurate with the degree of compliance with specifications.

To open up broad range of services of the CPSI to both individual work steps and the life cycle of products, the reconstruction of this product life cycles⁶ is chosen as a research area. It is demonstrated that the life cycles of processes and products are closely analogous. Like activities, a product passes through several intermediate states in its life cycle. Their detailed investigation forms another major part of the paper. It is outlined that the EPCglobal standard can be used to make radio frequency identification (RFID) events accessible for process mining, even inter-organizational process mining. As with the RFID events, some efforts are also made to reconstruct data in enterprise systems. This demonstrates that process mining is not restricted to process-aware information systems (e.g., workflow

⁶ Note that this life cycle is a succession of assembly and disassembly stages an individual product goes through – this in contrast to the phases market introduction, growth, maturity, saturation, and decline.

1. Introduction

management systems (WFMSs)) and to single-systems, but can be implemented more comprehensively, notably in supply chains. This has remarkable implications because it has a positive influence on the further development and deployment of supply chain analysis.

The successful use and integration of the designed generic components into the novel CPSI system demonstrate the validity and potential of the methodology.

3. *Implementations*, which realize highly required functions and services for the CPSI. The discussed theoretical concepts are generally complemented with practical implementations.

The first implementation addresses the shortcomings identified in measuring the compliance of processes with reference models having novel measurements. The main idea behind this implementation is to search for the maximum number of identical activities while preserving the characteristics of reference models. The categorization, analysis, and practical exploitation of vulnerability raise awareness for compliance problems and point to the compliance mechanisms, which should generally be treated as an integral part of the CPSI instead of just an add-on.

The second implementation supports the preparation of RFID events. An intriguing aspect of this algorithm is that it identifies a case and handles different types of events, including assembly and disassembly events, as a result of which focus shifts are managed.

The data preparation for process mining purposes further results in an application-specific algorithm that is specialized for responding to transactional conditions in an enterprise system, notably a customer relationship management (CRM) system.

The implementations described can be processed using the process mining framework (ProM) now. The design for these algorithms is based on open source software and uses off-the-shelf hardware. These prototypes probably need little or no adjustment for a huge class of IT-enabled processes.

The work on this topic is completed by augmenting it with the seamless integration of the implemented approaches into the overall improvement approach. The benefits of the designed solutions are clearly demonstrated by applying them to the case studies. The results show that the improvement system brings considerable gains in process quality and maturity and fills extant management gaps in IT instruments. As for non-functional requirements, no performance penalties for violating runtime constraints result from the implementations.

1.4. Outline

This section gives an overview of the structure of the following chapters.

Chapter 2 outlines the organizational aspects of managing business processes as well as the technological support for these processes of the CPSI perspective. Starting with a historical retrospective the role of process organizations is discussed from the view of

management theory and important terms relating to it are defined. Next, the focus is on the characteristics of current BPM approaches that provide the conceptual framework for the CPSI. The emphasis is then on the importance of business process modeling and the role of improvement in the business process management life cycle. The chapter continues with attention to reference models and process mining. The introduction of the technological support opens with a discussion of the role and value of IT before addressing IT-Business alignment, compliance and risk, and IT governance. Finally, it introduces the RFID technology and its related EPCglobal standard.

Chapter 3 explores new data sources for process mining. It begins by reviewing the literature. Next, it introduces the theory of case construction. It then cites challenges for making electronic product code information services (EPCIS) events and data from enterprise systems, such as in the SAP Business Suite, applicable to process mining. In view of these challenges, concepts for data preparation are designed and implemented that derive case identifiers from EPCIS events and enterprise data, thereby contributing as to applying process mining for supply chain analysis. The chapter concludes with a number of lessons as to the usability of RFID events and enterprise data.

Chapter 4 analyzes empirically the way process mining has been useful to the case studies' organizations and identifies the challenges they likely faced when applying process mining. This chapter initially both presents a set of use cases with which improvement issues in their real word setting are identified and advocates the case study methodology in business research. After that, the expediency of the concepts developed in the preceding chapter is demonstrated by employing them to the case studies. This chapter dwells as well on applying process mining to the well-known data source *workflows* in practice for process improvement. Finally, it summarizes the benefits and drawbacks and discusses the data preparation effort and its effect on process quality.

Chapter 5 gives attention to the development of a new CPSI approach, one integrating process mining and reference models. In addition to optimizing processes, the CPSI approach is also concerned with the most effective use of limited resources in terms of personnel, systems, and resources. It gathers state-of-the-art works on event log analysis, process-based controlling, and data warehouse concepts. A survey is made of the possibilities of quality management based on ITIL and process mining through an example of incident management, a central ITIL service operation process. The definition of the business process control framework starts from analyzing the need for a new framework, goes on to investigate the relevant entities, and proceeds to introduce the procedure model of IT service operation with respect to ITIL and process mining. Next, the approach is evaluated in practice by applying it to the use cases. The development of the approach rests on a single reference model; the compatibility of another reference model, COBIT, is demonstrated. Finally, both the conclusions reached during implementation in practice and future research directions are described.

Chapter 6 presents and evaluates the rationale and concept needed to establish a novel algorithm supporting the process improvement approach by measuring the degree to which a process model complies with a reference model. The first topic explored is a contemporary account of related work, mainly techniques to determine equivalence and applications to measure compliance. Based on the use cases, requirements are

1. Introduction

derived for determining compliance with reference models. This algorithm is presented along with details of the concept for the improvement evaluation and a corresponding implementation as a plug-in for ProM. The approach is based on comparison of process instances produced by a process model. Next, the approach is evaluated by measuring the compliance of the model currently used by the use cases' organizations with the ITIL reference model and by comparing the results with extant approaches. A feasibility study using a sample of event-driven process chains (EPCs) of the SAP reference model deals with whether or not the computation times of the algorithm are acceptable, and then whether the algorithm is practicable.

The dissertation concludes in Chapter 7 with a summary of the activities and results. It illustrates the current status of processes in the management domain. Based on this description, it gives an outlook for potential enhancements as well as the likely relevance and applicability of the approaches. The implications of this paper for guidelines and management for the continually improving processes and respective tool support are discussed at length.

In addition, five appendices are provided. Appendix A comprises a questionnaire analyzing the overall view of practitioners from industry on IT management activities in light of IT's contribution to business process improvement. Appendix B comprises the concomitant analysis of the questionnaire. Appendix C comprises process models derived throughout. Appendix D comprises a detailed presentation of two mining algorithms, namely the *α -algorithm* and the *Heuristics miner*. Appendix E comprises the Supply Chain Operations Reference-Model (SCOR) (deliver make-to-order (D2)) scenario, which represents the delivery strategy in which products are delivered only in response to a customer order.

2. Process Organizations

The following chapter outlines the organizational aspects of managing business processes as well as their technological support from a perspective of the continuous process improvement (CPI). No claim is made that the details are comprehensive. However, a succinct conceptual description is proffered to provide a general understanding of both the BPM and the IT management.

Section 2.1 presents a historical retrospective of process organizations⁷ and provides definitions of BPM. Section 2.2 focuses on the characteristics of current BPM approaches that provide the conceptual framework and techniques for the CPSI. Section 2.3 introduces various approaches to modeling that can be used for the representation of a business process model. Section 2.4 provides the background information for reference models. Section 2.5 gives an overview of process mining. This chapter ends in Section 2.6 with the technological support for process organizations.

2.1. Business Process Management

This section starts with a historical retrospective and discusses the role of process organizations from a management theory perspective. Given the fact that the evolution of business management is intermediately related with IT developments, the retrospective recognizes the progress in business and in IT. Finally, core concepts of BPM are provided.

2.1.1. Background

The twentieth century wrought more scientific, industrial, and technological changes than all prior human history (Mansir and Schacht, 1989, Chapter 1, p. 6). These changes have progressively altered the landscape and nature of process and IT organizations (Dumas et al., 2005). It is therefore interesting to place the evolution of organizational structures and the underlying IT developments in a historical perspective.

Functional Separation

Obviously, the works of *Henri Fayol* and *Frederick Taylor* enormously influenced twentieth century organizational theory. Both Fayol and Taylor highlighted the need for

⁷ Process organizations are organizations viewed from a process perspective (Leon, 2007, p. 349). It operates across functional boundaries to create an environment noted for excellence in efficiency, effectiveness, and customer satisfaction. Bangemann (2005, pp. 119,120) presents distinctive criteria in terms of process and functional organizations.

2. Process Organizations

coordination, though from different perspectives. Fayol's research on managerial structure of an organization is a seminal work of classical organization theory. He proposed five primary functions of management: planning, organizing, commanding, coordinating, and controlling. In addition, Fayol (1966, p. 20) introduced fourteen principles of management; among them are the division and specialization of labor, and managerial coordination. As specialization increased output by making employees more efficient, the need for coordination between the individual tasks increased. Taylor (1911, pp. 61–62), however, was deeply concerned with the operational execution of tasks. He devised four principles based on the assumption that standardization of methods, adoption of best implements and working conditions, and cooperation all contribute to the efficient execution of tasks. Until the 1980's a great many organizations embraced his principles. (There are organizations today that still use them (zur Muehlen, 2004).) *Henry Ford (1926, p. 105)* exemplified this, applying as he did Taylor's scientific management principles to increase the productivity of the individual worker in his automobile factories in the 1920's.

From an IT perspective, the 1970's and 1980's were dominated by data-driven approaches. The emphasis on IT was in storing and retrieving information from database management systems or material requirements planning (MRP) systems. Data modeling rather than process modeling was the starting point for building information systems. As a result, ironically business processes were often structured to accommodate underlying IS (van der Aalst et al., 2003a), thereby confounding the original business model.

From a managerial perspective, separate functional business areas within an organization were suitable for the market conditions for many decades until the 1970's; supply strategies drove management thinking because markets were stable, consumer behavior predictable, and product life cycles long (zur Muehlen, 2004, p. 45). The 1980's, though still dominated by functional approaches to organizational design, saw optimization potential striven for the use of cross-functional improvement. Subsequently, lagging strengthened the historically unenviable position of consumers – as a result, price and quality of products and services became a crucial selling point in many markets. Among other things, growing consumer cost consciousness, shorter product life cycles, and increasing market segmentation, led both industry and academia to seek structures better suited to changing market conditions and enterprise infrastructure (zur Muehlen, 2004). Process orientation has therefore become mainstream thinking for organizational research.

Process Orientation

The importance of the alignment of organizational structures with business processes was discussed in the literature since the 1930's. Authors such as *Nordsieck (1934)*, *Henning (1934)*, and *Chapple and Sayles (1961)* are among the prescient proponents of process orientation. Despite early interest in academia, the process-oriented organizations did not emerge substantially in industry until the early 1990's, when workflow technology, such as FlowMark and Staffware, emerged to support business processes' execution. Workflow management systems manage and execute entire business processes involving

personnel, applications, and information sources according to defined business process models, currently referred to as process-aware information systems (PAISs) (Dumas et al., 2005, p. 7).

The formal description of business processes was aided by business process modeling tools, exemplified by Protos and Architecture of Integrated Information Systems (ARIS); they arose virtually contemporaneously with workflow technology in the early and mid 1990's. Workflow technology unquestionably helps organizations in the transition to a process-oriented organization (zur Muehlen, 2004) – but the management concepts of proponents of a process-oriented view of business management like *Porter (1985)*, *Davenport (1993)*, *Harrington (1991)*, and most notably *Hammer and Champy (1993)*, contributed likewise to the fact that business processes have become the focal point of organizational design. These management approaches (e.g., BPR and CPI, mainly proposed also in the early 1990's and the mid 1990's, concern the applications of IT (Reijers, 2005). Indeed, the consequences for IT itself and for the broader society have been transformational (Österle, 1995, p. 12); and the once overwhelmingly industrial society has changed into an informational society. Take the development of Electronic Data Interchange (EDI). This technological innovation enabled the electronic communication of information between supply chain parties though up to the late 1990's, the improvement and automation of intra-organizational business processes remained the predominant focus of BPM (Dayal et al., 2001). Since then technological innovations (e.g., Internet technology, Service-Oriented Architecture (SOA), web services) have led to a number of changes that permit greater freedom on the part of business organizations. These technologies enable organizations to distribute information and resources swiftly or instantaneously by space and time (Österle, 1995, p. 12). From the corporate management perspective, this means a complete rethinking of all aspects of business (Österle, 1995, p. 12). The adoption of concepts, such as real enterprise, business on demand, and adaptive enterprise, has resulted increasingly in the expectation that IT will contribute to business success both more flexibly and directly in a way that can be gauged. Therefore IT could no longer afford to look exclusively at its internal organization – it had to consider the quality of the services more broadly, focusing on the relationship with customers (van Bon, 2002). As a result, ITSM has evolved commensurately as the use of service management grew in time. The 1990's witnessed the rise of ITSM employing best practices, salient examples of which include ITIL and COBIT. These best practices promote both a purposeful development of application systems and the management of the IT business. ITSM became a common term in the mid 1990's (Office of Government Commerce, 2007).

Under the keyword “IT governance” (Weill and Ross, 2004; IT Governance Institute, 2003, 2007) various concepts are currently developed and discussed; among them are IT-Business alignment, risk management, resource management, IT value delivery, and performance measurement. These new concepts, about which more in due course, call for an extension of the tasks of the IT management and the ITSM.

2. Process Organizations

2.1.2. Definitions of Business Process Management

This section provides broad definitions of BPM against the background of organization theory. For the discussion of BPM, this section offers a general definition of the term *business process*. Detailed explanations of the process terms are given in Harrington (1991, p. 9) and Johansson et al. (1993, p. 53).

In 1934 Nordsieck (1934, pp. 27–29) described a process as a sequence of activities producing an output, and an activity as the smallest discernible unit of work performed by a subject. Davenport (1993, p. 5) adopted a similar view, emphasizing that the unit of work was a structured and measured set of activities across space and time.

“A process is a structured, measured set of activities designed to produce a specified output for a particular customer or market.”

The measurability is a prerequisite for the BPM; as Harrington (1991, p. 164) has phrased it, “If you cannot measure it, you cannot control it. And if you cannot control it, you cannot manage it”. Dumas et al. (2005, p. 157) devote their description precisely to the ordering of activities, stating that activities of a business process are in logical relation to each other – alternative and parallel executions are possible. A number of authors differentiate between tasks and activities. Dumas et al., for example, call the execution of a task an activity – consisting of tasks in which each possible run of the process consists of activities referring to the tasks. They define a process as follows:

“A process consists of tasks which have to be executed. These tasks can be in some order (sequentially), stating that one task can only be executed after the execution of another task is finished. If two tasks are not ordered, then they can be executed concurrently. Tasks can also be alternative, that is, if one task is executed, then the other task is not executed and vice-versa. Tasks can be executed more than once in general. A process can be in different states. A process starts with an initial state (which is not necessarily unique) and might end with a final state (which is also not necessarily unique). Usually, it passes through several intermediate states” (Dumas et al., 2005, p. 157).

A business process can be viewed from various perspectives, namely data, function, organization, and control flow – Section 2.5.2 will discuss them. The execution of individual business processes is also referred to as business process instances, or merely “instances”. The execution of business processes is supported by information systems (Allweyer and Scheer, 1995; Österle, 1995, p. 18), among which are Enterprise Resource Planning (ERP) systems, CRM systems, and workflow management systems.

Since organizations face the challenge of managing business processes (Mansir and Schacht, 1989, Chapter 1, p. 6), the term BPM requires a careful definition. The Australian BPM Community of Practice (2009) proposed the following:

“Business process management is a structured, coherent and consistent way of understanding, documenting, modeling, analyzing, simulating, executing, and continuously changing business processes and all involved resources in the light of their contribution to business improvement.”

Four phases in this definition are especially noteworthy: structured, continuously changing, process, and business improvement. Just as this definition stresses the “structural” character of BPM, various authors have arranged the management activities in the form of a life cycle as part of the overall structure of BPM with the aim to improve outcome: BPM (zur Muehlen, 2004), BPR (Dumas et al., 2005), CPI (Neumann et al., 2003, pp. 239–244), and WFMS (Rolles, 1998, p. 128; Heilmann, 1994, p. 14). Since singular attention is paid to processes, note that additional life cycle models focusing on objects – rather than on processes – are in use; consider product development life cycles, customer resource life cycles, and software development life cycles.

The phrase “continuously changing” denotes change’s inevitability, its ever-increasing pace (Mansir and Schacht, 1989, Chapter 1, p. 6), and that organizations therefore face the challenge of providing flexible business processes, which themselves must be adapted to protean environments (Mansir and Schacht, 1989, Chapter 2, p. 1). Regev et al. (2006) see business process flexibility as the capability to implement changes in the business process by changing *only* those parts that need to be changed and by keeping other parts stable. The third phrase refers to “processes” – in order to achieve business improvement, it is necessary for an organization to subordinate much or all to its primary business processes (Dumas et al., 2005, p. 90). The phrase “business improvement” signals the indispensability of steady improvements for organizations to achieve efficient and effective execution of business processes.

One addition to the definition is needed: the term “controlling” as a stage through which a business process goes. Zur Muehlen (2004, p. 3) cogently argues that continuous maintenance and control of the business processes is required to obtain continuous benefits from a process-oriented organization.

The integration of information systems is characteristic of BPM approaches (Reijers, 2005).

2.2. Business Process Management Concepts

To achieve business improvement, various BPM concepts have been developed: business process engineering (BPE), business process reengineering, and continuous process improvement – these will be briefly discussed.

2.2.1. Business Process Engineering

Business process engineering is associated with decisions at all levels in the structuring of an organization (Österle, 1995, p. 15). Decisions are made as to business strategy, process level, and IS level. Business strategy covers corporate policy decisions concerning the goals and functions of an organization (Kagermann and Österle, 2006). The definition includes the role of the organization in the supply chain and its contribution to the creation of value. Core competencies, alliances, and logistics are specified as well (Winter, 2003). The process level in turn determines the following: organizational units, process outputs, subprocesses and organizational activities as well as the most important transactions in the database necessary for the realization of the defined strategies

2. Process Organizations

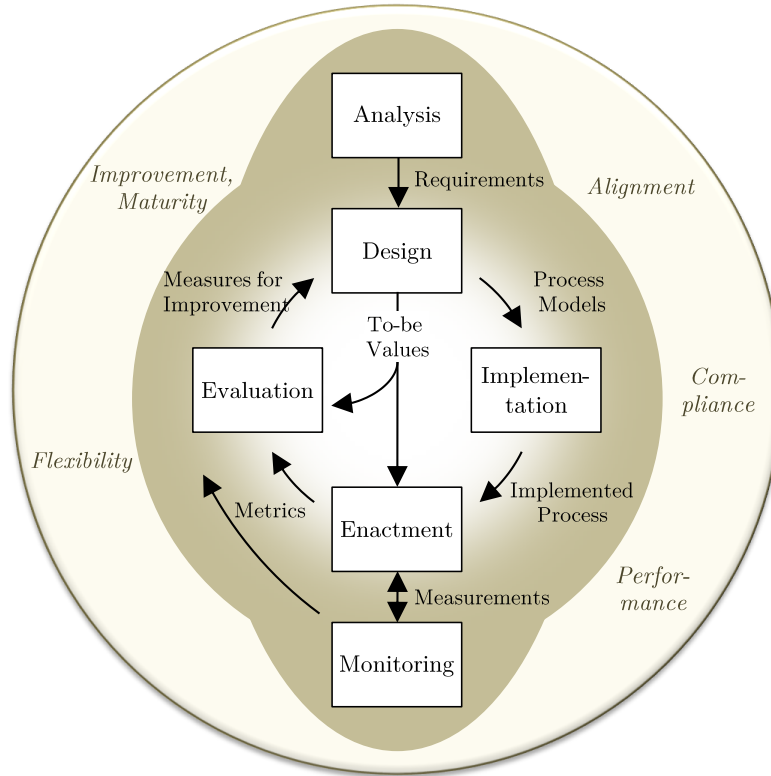


Figure 2.1.: Life cycle of business process management

(Österle, 1995, p. 16). Finally, the IS level specifies the computerized information processing – this can be seen from such information as responsibilities, computer screen masks, rights of access, and dialog flows (Österle, 1995, p. 16). There are two types of business process engineering: forward engineering and reverse engineering. Forward engineering is the well-known process of moving from a high-level abstraction – that is, the business strategy in terms of business engineering – to the implementation. Reverse engineering is a bottom-up approach that analyzes a finished product, a system, or an end result of a work process to determine how it was realized.

Similarly, the business process reverse engineering (BPRE) focuses on the analysis of an IS with the objective of recovering its design or specification. These inductive analyses are naturally far less subjective than assessments based on interviews, whether they focus on a key person or on numerous employees. The quality of the process models extracted by BPRE approaches is higher than process models prepared conventionally (Herbst and Karagiannis, 1998). Reverse engineering differs from re-engineering – the former is used to derive the design or specification of a system from the available input – the latter is used to produce a new system. Reverse engineering substantially contributes to a better understanding of a system, which is often part of a re-engineering process.

Considering BPM life cycles in detail reveals that various phases support the management of business processes. Each phase employs exactly the same management activities. The BPM life cycle presented in zur Muehlen (2004, pp. 82–87) is here adopted for two reasons. First, he took supplementary information from the business process life cycle by Neumann et al. (2003) who themselves had strongly emphasized the CPI. Second, his life cycle is related to the cycles for workflow modeling – therefore it is for an effective and efficient IT support of business processes, ensuring as it does ample integration of business processes and IT. The BPM life cycle is shown in Figure 2.1. The phases analysis, design, implementation, enactment, monitoring, and evaluation appear. They are arranged in a cyclical structure demonstrating their logical relations. The phases’ purposes are denoted in the outer circle.

Analysis. The BPM life cycle begins by analyzing the business processes necessary to deliver results in accordance with the organization’s strategy. The analysis considers both the organizational and the technical environment of processes and culminates in a set of requirements. Techniques, such as fishbone diagramming, cognitive mapping, and Pareto diagramming, ensure exhaustive analysis.

Design. Once the requirements have been specified, the design phase follows in which the overall process structure is engineered, validated, and finally represented by business process models. Business process modeling techniques and verification as well as simulation techniques are used – business process modeling is the most important one (Weske, 2007, p. 12) as it formalizes the business process description. This formalization is maintained by using a particular business process modeling notation, a topic that will be fully addressed in Section 2.3. In addition to the process definitions, to-be values are set. They are used to measure the success of the improvements as input for evaluation and enactment activities. With respect to measurement Reijers (2005) distinguishes four major performance aspects that can be targeted for improvement: time, cost, quality, and flexibility. The design activities align the business process with the requirements of the business.

Implementation. At this point in the life cycle, the infrastructure for business process support is designed and the solution is integrated with surrounding IS. Implementation ensures compliance of the business processes with specifications (e.g., customers’ requirements and regulations), tests the performance of the business processes, and releases them for operation. The completed process models are input for the process implementation phase.

Enactment. During the process enactment phase, individual instances are executed within the infrastructure implemented (van der Aalst et al., 2003a). Process metrics about the performance are collected. Enactment measures and assesses the performance of the instances using predefined criteria and to-be values. Anomalies can be quickly detected (Dumas et al., 2005). The enactment of business processes ideally is continuous.

2. Process Organizations

Monitoring. Simultaneously, process monitoring occurs. The term monitoring denotes all activities that are associated with the continuous supervision of to-be values. It can therefore be used for assessing both the performance of the process management system itself and measures such as idle time of resources and delay of pending activities.

Evaluation. The process evaluation phase completes the process management cycle. The execution of instances is analyzed from a perspective retrospectively based on protocols. Examples are audit trails and event logs. Measurements are contrasted with the results sought – deviations that prevent the attainment of objectives are revealed. This analysis can result in adjustments to process structure, which can be tested during simulation with regard to their impact on process performance (zur Muehlen, 2004, pp. 82–87). Further techniques, such as activity-based costing and time-motion study, are available to support the process evaluation. Obviously, evaluation activities are introduced for the purpose of improving a process and for determining its maturity – at the same time they cannot be employed rigidly, given the dynamic reality of the business and its goals.

2.2.2. Business Process Reengineering

A premise of the business process reengineering is the fundamental revaluation of business processes – both the changes and their outcomes must be far reaching – for the most effective and efficient possible business process structure to be achieved. The BPR is a means to break with current ways of business by supplanting present processes with utterly new ones (Dumas et al., 2005, p. 90). Hammer and Champy (1993) define BPR as

“[...] the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical contemporary measures of performance.”

The BPR life cycle proposed by Dumas et al. (2005) is designed in four phases: diagnosis, redesign, reconstruction, and operation. The life cycle starts with the diagnosis phase in which an analysis is made of problems that arise or are in processes. The redesign phase establishes an entirely new description of the processes. To support the activity of redesigning, creativity techniques like out-of-box-thinking, affinity diagramming, and the Delphi method are available (Reijers, 2005). The classification of Kettinger and Teng (1997) neatly describes methods, approaches, and guidelines that deal with the BPR. Interestingly, Reijers (2005) also employs reference models as a technique for the BPR. Consideration of these is left until Section 2.4. During the reconstruction phase, a new system is built to support the processes previously identified. Finally, the system begins operation.

2.2.3. Continuous Process Improvement

The continuous process improvement like the BPR is a process-centric BPM approach. The CPI is predicated on the fact that an organization must continuously adapt to its changing environment (Mansir and Schacht, 1989, Chapter 2, p. 1) lest performance and quality improvement suffer (Mansir and Schacht, 1989, Chapter 1, p. 5). Rather than striving for radical change, one using the CPI produces incremental improvement toward set goals and swift adaptation to changing conditions. In this context the CPI takes current business processes as a starting point, they are gradually refined (Reijers, 2005) – this in contrast to the BPR, the use of which presumably brings about entirely new business processes. The CPI engages the whole organization in improvement activities – it compels senior executives to foster an environment conducive to rapid modification and enhancement (Mansir and Schacht, 1989, Chapter 1, p. 8).

Since the BPM life cycle integrates the fundamental principles of Neumann et al. (2003), similarities and differences between BPR and CPI follow.

Unlike the modifications brought about by BPR elements, those of CPI are generally not radical and occur more frequently. The concepts BPR and CPI resemble one another – each is committed to quality. Their definitions range from narrow ones, such as “conforming to specifications”, to broad and perhaps abstract ones, such as an intangible but intrinsic goodness (Mansir and Schacht, 1989, Chapter 3, p. 5). For example, total quality management (TQM), Six Sigma, and Kaizen are approaches organization may employ to support the BPR or the CPI.

2.3. Process Modeling

This section introduces different approaches to modeling that can be used for the representation of business processes.

A broad range of notations provides constructs and rules for the description of business processes – from these notations Petri net, EPC, and business process modeling notation (BPMN) are selected. The selection illustrates differences and commonalities without going into mapping details. The primary concern is with those concepts of modeling that are applicable to processes. This means that other types of modeling done by organizations for business purposes are beyond this paper’s scope. Hence, neither entity relationship model (ERM) (Chen, 1976) for the representation of data structures, organization charts for organizational structures and resources (Scheer, 1997, pp. 23–30), information models (Scheer, 1997, pp. 690–697), nor business rules (de Beer, 2004) are considered. The business process execution language (BPEL) (OASIS, 2007) has been receiving increasing attention for specifying business processes; but this modeling language is excluded as well as it only addresses the execution of business processes.

The meaning of a model has to be fixed before suitable notations for modeling processes can be introduced. The term *model* is pertinent in numerous contexts. In business informatics, models are seen as simplifying a part of the reality, and they can be seen purpose-related representations of it (Hars, 1994).

2. Process Organizations

The main aim of business process modeling is to provide means for a precise description of business processes using notations with formal semantics (Dumas et al., 2005). These models need such foundation for several reasons. First, formal models are unambiguous so that all concerned assign the same meaning to them (van der Aalst et al., 2003a), and they may also be used for documenting the knowledge reposing in the various components of a process organization, among which are information systems and employees. Second, the graphical representation of a formal model facilitates the understanding and the transparency of process organizations so that different actors can communicate process structures efficiently (Weske, 2007, p. 11). Third, formal modeling notations may have associated analysis techniques that can monitor and control the behavior and the characteristics of a business process model (van der Aalst et al., 2003a). Such analyses can also be used to improve and refine processes (van der Aalst et al., 2003a). Process modeling can be closely linked to BPM concepts both for design and redesign. Finally, a formal model can be employed as a starting point for building an IS.

The following explains key issues in modeling business processes with Petri nets, extended event-driven process chains (eEPCs), and BPMN. A simple example from an order processing follows. The process is always initiated by a customer's order. As can be seen in Figure 2.2, all products are either available, that is, they are retrievable from stock or unavailable, meaning they have yet to be produced. As soon as the order has been released, the customer is billed and the order is shipped; billing and shipping may occur in parallel (concurrently), or either one can precede the other. After both are done, the order is closed.

2.3.1. Petri Net

Unlike other approaches, Petri nets do not only have a uniform graphical notation but also a sound mathematical foundation (Dumas et al., 2005, pp. 147 ff.). That is exactly why Petri nets are popular in science and industry. Formal semantics are a prerequisite for simulation and analysis methods. For an introduction to Petri net theory, see Murata (1989), Reisig and Rozenberg (1998), or Peterson (1981).

Since the development of Petri nets by *Carl Petri (1962)*, they have been supported by tools from both commercial vendors and academic institutions. Petri nets are sometimes tailored for specific application domains or different abstraction levels as can be seen by colored Petri nets (CPNs) and workflow nets. In particular, the nets are useful for modeling systems in which behavior is dominated by the flow of information, objects, and control. They are thus well-suited to model business processes. An apt definition of a Petri net based on its graphical appearance is provided by Dumas et al. (2005, p. 152):

“A Petri net is a directed graph with two different types of nodes: places, represented by circles, and transitions, represented by rectangles. Petri nets are bipartite, that is, no arc connects two places or two transitions. Nodes and arcs can have various annotations.”

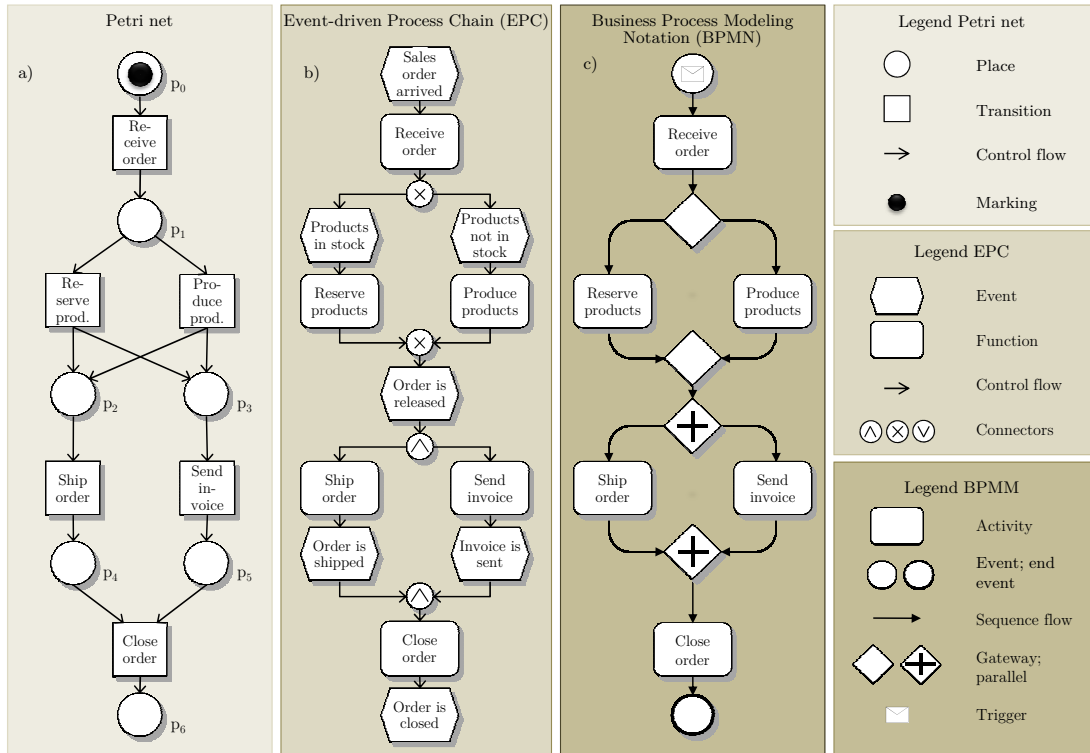


Figure 2.2.: Some examples of modeling notations

However, to define formally the structurally restricted Petri net class workflow net (WF-net) in Section 6.4.1, the mathematical definition of a Petri net will be used. Now Dumas et al. (2005, pp. 153, 158) have stated:

“A Petri net consists of two disjoint sets S (places) and T (transitions) and a binary relation F . [...] Usually, a Petri net can be equipped with markings, where a marking is a mapping from the set of places to some domain. [...] Pre- and postconditions of tasks are modeled by places, which are in the post- and presets of the respective transitions. The idea is that for ordered tasks, the occurrence of the first transition produces a token on the place, whereas the second transition is only enabled after this token is produced, and it consumes the token.”

Such a representation can be in- and output to Petri net tools. The Petri Net Markup Language (PNML), for example, is based on the Extensible Markup Language (XML) (Dumas et al., 2005, p. 154).

In the example, the Petri net model in Figure 2.2 (model a) includes an initial marking (or token), as can be seen in the case for the place *Receive order*. At this stage, only the transition representing activity *Receive order* is enabled. Its execution leads to a new state, one in which the place p_1 is now marked. This marking enables the transitions

2. Process Organizations

representing the activities *Reserve products* and *Produce products*. They compete for the marking because they share a common place in their presets – there is a choice (XOR-split) between the transitions. The corresponding XOR-join is modeled by common postsets in the places p_2 and p_3 . The occurrence of one transition disables the other transition and it distributes markings to two places so that subsequent transitions *Ship order* and *Send invoice* are enabled concurrently. This occurs because their preconditions are disjoint (AND-split). The final marking can only be generated when both the right activity and the left activity have finished (AND-join).

2.3.2. Event-Driven Process Chain

The EPC was developed in 1992 in a research and development project in which SAP AG was engaged. It is the core component of SAP R/3 modeling concepts for business engineering and customization (Dumas et al., 2005, p. 119). It is also part of the method family ARIS, which forms the methodical basis of the BPM after IDS Scheer AG (2007). The basic EPC notation has been extended with symbols corresponding to various aspects of business modeling, among which are data elements, and organizational and resource views. This has led to what is known as the eEPC. As the extended version includes the basic EPC notation, no distinction is made between the original one and its extensions⁸.

The eEPC captures the control flow of a process in terms of the temporal and logical dependencies of activities (Keller and Teufel, 1998a). Figure 2.2 (model b) shows the eEPC example. The eEPC notation offers functions to represent activities, events describing pre- and postconditions of functions, and three types of logical connectors linking functions and events: AND, OR, and XOR. These building blocks are connected with control flow arcs. An event is represented by a hexagon. An activity is represented by a rectangle with rounded edges. Connectors are represented by circles. They have either multiple incoming arcs and one outgoing arc (join connectors) or one incoming arc and multiple outgoing arcs (split connectors) (van Dongen et al., 2008). The informal semantics of an eEPC can be described as follows: The AND-split occurs at a point in the process in which activities can be executed in parallel. The activities can be executed simultaneously or in any order. They are converged again via synchronization (AND-join). The exclusive choice (XOR-split) is a point in the process from which one of several options is chosen. With the help of a simple merge (XOR-join), two or more alternative activities come together without synchronization. The exclusive choice by definition means that none of the alternative activities is executed in parallel (Russell et al., 2006). An eEPC forms a bipartite graph in which no arc connects two events or two functions.

Dumas et al. (2005) argue that sacrificing a strong formal framework for the sake of simplification has led to the widespread acceptance of extended event-driven process chains among industry (Dumas et al., 2005). However, various approaches to eEPC semantics formalization exist. They are primarily based on the concept of mapping

⁸ Note, that the term eEPC is always used to refer to both the basis and the extended notation because the acronym “EPC” also stands for electronic product code.

eEPC elements to Petri net fundamentals (Chen and Scheer, 1994; van der Aalst, 1999; Dehnert and van der Aalst, 2004). The ARIS Markup Language (AML), for example, is based on XML. Such a representation can be in- and output to Petri net tools.

2.3.3. Business Process Management Notation

The BPMN is a standardized graphical notation for creating business processes. It was developed by the Business Process Management Initiative (BPMI) and is now maintained by the Object Management Group (OMG). It was released in 2004.

The notation was specifically designed to provide a lucid notation and to handle the complexity intrinsic to business processes. The four categories of elements are flow objects, connecting objects, swimlanes, and artifacts.

Flow objects are graphical elements defining the behavior of a business process. There are three flow objects: events, activities, and gateways. Events are empty circles to allow internal markers to differentiate between a number of triggers or results. See model Figure 2.2 (model c). The first event is marked by an envelope indicating that a written letter has initiated the order. Activities are represented by rounded rectangles. A gateway, represented by a diamond, is used to control the divergence and convergence of sequence flow – it determines branching, forking, merging, and joining of paths. Flow objects are linked by the connecting objects sequence flow and message flow. A sequence flow is used to show the order in which activities are performed. A message flow is used to show the flow of messages between participants. The third flow object, called association, is used to relate information to flow objects. Swimlanes (pools and lanes) allow the grouping of primary modeling elements. Artifacts, comprised of data object, group, and annotation, are used to provide additional information about the process.

Seeking to ensure that XML designed for the execution of business processes can be visualized with a business-oriented notation (Object Management Group, 2006), there has been much activity in the past few years for the development of web service-based XML execution languages. The BPEL, for example, provides formal mechanisms for the definition of business processes specified in the BPMN.

2.4. Process Engineering and Reengineering with Reference Models

Reference models have been receiving increasing attention for design and redesign approaches in the literature in the last twenty-five years (Reijers, 2005). Organizations have begun, albeit only in the past decade or so, to take notice of them (IT Governance Institute, 2008a, p. 36). Cognizance of COBIT, for example, has surpassed 50 % percent – adoption has steadied up at around 30 % percent (IT Governance Institute, 2008a, p. 36). This section defines reference models and explains the motivation behind their use. Section 2.4.2 lists four categories of reference models and accompanying examples.

2. Process Organizations

2.4.1. Definition of Reference Models

Reference models were discussed early in business administration literature. Nordsieck (1931, p. 160) described so-called “task structuring plans” (*Aufgabengliederungspläne*) as follows:

“Usually, a task structuring plan already has a relatively universal character because it is created according to logical principles, that is, it is not only valid for the company being studied but rather – with a few changes – for companies with similar aims and the same branch of trade.”

Despite this early reference, the term reference model did not firmly establish itself in the literature until the end of the 1980’s, at which time reference model’s definitions resemble from decades earlier. Both Scheer (1989) and Fong and Jefferson (1986) asserted that reference models are universal and that they are commendable, two noteworthy facts.

Many authors, such as vom Brocke (2003, p. 31) and Hars (1994), deemed universality to be an intrinsic attribute of a reference model. *Alexander Hars* particularly illustrated this, writing that the abstraction from individual characteristics provides the essentials for the derivation of a specific model from a reference model (Hars, 1994, p. 15). Bear in mind that the universality of a reference model does not automatically mean that it has a claim to absoluteness (Thomas, 2006). A reference model can only make that claim with regard to a certain category of applications, as in a functional area. The literature continues to note reference models’ “recommendation character” (Scheer et al., 1994, p. 92; vom Brocke, 2003, p. 32). Reijers (2005), for example, proposed the following definition:

“A reference model is often seen as some sort of pattern, expressing the best way to treat a particular problem, which can be replicated in a similar situation or setting.”

It is emphasized that reference models offer an exemplary description from which enterprise specific models can be derived (Scheer et al., 1994, p. 92), these models taking on the role of to-be models. Because reference models are commendable, they are also referred to as best practices or good practices.

Patently, reference models, being both universal and commendable, ought to be adapted (Reijers, 2005).

Reference models are useful for many reasons. First, they optimize the design of processes. Benefits with regard to process design are illustrated by discussing four factors: quality, time, cost, and risk. Reference models raise levels of *quality* because they are generally developed over a long period and often capture experts’ insight (van der Aalst et al., 2006). The design *time* of process models can be sped up (Scheer, 1994) due to their reusability and quality. Adhering to reference models reduces data management and integration *costs* – this in turn mitigates investment *risks*. Risk reduction also occurs when reference models facilitate compliance with industry regulations (IT Governance

Institute, 2007). Second, reference models are an excellent means of creating a link between business needs and IT implementation (van der Aalst et al., 2006) – and they serve as a communication means in the organization. Finally, reference models can provide needed guidance in process improvement. The nature of the standard allows meaningful comparison with extant processes and provides an orientation for improvement activities.

2.4.2. Domains

Reference models have been collected and applied in various domains, such as software development, IT governance, and software development (Rosemann and van der Aalst, 2007; Reijers, 2005). As discussed earlier, reference models are only valid in a certain context. Therefore, numerous authors differentiate reference models along their scope (e.g., functional areas covered), their granularities (i.e., number of levels of decomposition detail), or the views underlying a model (e.g., processes, data, objects, and organization) (van der Aalst et al., 2006). A comprehensive differentiation based upon domains can be found in Pesic and van der Aalst (2005).

The view on reference models herein is based on functional areas according to the categorization of van Bon and Verheijen (2007), who proposed four functions: information management, governance, quality management, and quality improvement. This categorization comprises reference models (business process reference models as well as IT process frameworks) that focus on capturing process-oriented best practices. Table 2.1 illustrates this. *Information management* reference models concentrate on efficient means of performing and organizing certain aspects of management, such as requirements, procurement, and service delivery. *Governance* reference models structure the processes in terms of responsibilities, controls, and organization. Reference models that emphasize quality standards applied to specific domains (e.g., services, security, and development) are called *quality management* reference models. Reference models that focus on the improvement of processes (performance or otherwise) without dealing at all with how-to aspects referred to as *quality improvement* reference models.

The list is not comprehensive. A first point is that several reference models contain material that could be allocated to more than one area, bearing in mind that the reference models are categorized by their primary functions. A second point is the process orientation of this paper. Since configurable reference models, such as the SAP Solution Manager (SAP AG, 2008), focus on the support for configuration and implementation of an IS rather than dealing with processes – these reference models are also not discussed.

2.4.3. Reference Models

SCOR, COBIT, and ITIL are selected from the domains introduced above – these models are widespread both in academia and industry to improve processes. This section summarizes the origin of the selected reference models, highlights their characteristics, and analyzes their role in process improvement and in compliance. This compliance is chosen as a source of process improvement because design and execution of processes might be able to take advantage of all the benefits discussed earlier. Note that some reference

2. Process Organizations

models use the term conformance rather than compliance. The in-depth description of IT frameworks of van Bon and Verheijen (2007) might also facilitate comprehending reference models.

Supply Chain Operations Reference-Model

SCOR was developed by the Supply Chain Operations Council (SCC) as the industry standard for supply chain management. The SCC is a nonprofit organization open to all companies that are concerned with applying supply chain management. Since its inception in 1996, SCOR has provided reference models that fuse business processes and technology features into a unified structure to support communication among supply chain parties and to improve the design and operation of supply chains (Supply Chain Operations Council, 2006). At this writing, the current version is SCOR 9.0.

The reference model is organized in three levels of decomposition: management processes, process strategies, and process activities (Supply Chain Operations Council, 2006, p. 7). The top level opens from a management perspective incorporating the five primary management processes: *plan*, *source*, *make*, *deliver*, and *return*. In addition to these basic processes, SCOR distinguishes three process types for each of the manage-

Table 2.1.: Categorization of reference models

Category	Reference Model
Information management	ITIL ⁹ , SCOR ¹⁰ , BiSL ¹¹ , ASL ¹² , eTOM ¹³ , TOGAF ¹⁴
Governance	COBIT ¹⁵ , COSO ¹⁶ , Val IT ¹⁷ , M_O_R ¹⁸ , ISO ¹⁹ 38500
Quality management	TickIT ²⁰ , ISO 9000, ISO 27001, ISO 20000, EFQM ²¹ , PCF ²²
Quality improvement	CMMI ²³ , BPMM ²⁴ , ITS CMM ²⁵

⁹ www.itil.org

¹⁰ www.scor.com

¹¹ www.aslbisfoundation.org

¹² www.aslbisfoundation.org

¹³ www.tmforum.org

¹⁴ www.opengroup.org/togaf

¹⁵ www.isaca.org/cobit

¹⁶ www.coso.org

¹⁷ www.isaca.org/Val_IT

¹⁸ www.mor-officialsite.com

¹⁹ www.iso.org

²⁰ www.tickit.org

²¹ www.efqm.org

²² www.apqc.org/pcf

²³ www.sei.cmu.edu/cmmi

²⁴ www.omg.org/spec/BPMM/

²⁵ www.itservicecmm.org

ment processes: *planning*, *execution*, and *enable*. The top level is refined on the second level according to three different strategies: *stock*, *order*, and *engineer*. Supply chain parties are free to choose from these strategies independently of each other. A supplier, for example, can use the deliver make-to-stock process strategy while the producer chooses the deliver make-to-order (M2) process strategy. Section 4.1.1 provides a comprehensive description of the M2 strategy. The third level decomposes each process strategy into activities in implementation. This level of detail specifies inputs and outputs of each of these activities, process performance metrics, benchmarking, and technology capabilities required to support the processes.

SCOR has particular significance for process improvement because it offers an in-depth description of activities involved in a process in conjunction with metrics and best practices. The use of process development methodologies and metrics for quantitative benchmarking, as well as the use of best practices for qualitative benchmarking permit the identification of opportunities for process improvement and facilitate their realization. However, the model is silent about persons and organizational elements that perform the activities.

SCOR is clearly indispensable to compliance – indeed, it requires organizations to be compliant with steps given in the management processes (Supply Chain Operations Council, 2006). Subjects of compliance are defined in the processes of the above type *enable* (such as EP.8 *Enable Plan* Regulatory Requirements and Compliance and ES.8 *Enable Import/Export* Requirements) and cover appropriate metrics and outputs for complying with regulatory documentations and process standards set by external entities, such as government and trade officials. The element EP.8 is concerned with laws and regulations and ES.8 deals with import and export regulation documentation.

Control Objectives for Information and Related Technology

The IT Governance Institute (ITGI)²⁶ developed COBIT. The first version was released in 1994.

As an open framework for IT, in its fourth version, COBIT 4.1 offers a set of generally accepted management activities and controls that provide guidance both in maximizing the benefits derived through the use of IT and in developing appropriate IT governance and control within an organization. COBIT organizes the IT activities in thirty-four processes (IT Governance Institute, 2007). As with structure, each process description follows the same approach. This description begins with a specification of its expected outcomes. It continues with a definition of the minimum controls to be considered in dealing with that particular process risk. The concept of measuring the performance is primarily based on the IT Balanced Scorecard (Kaplan and Norton, 1992), which is integrated in COBIT. Each process description closes with a maturity model describing an evolutionary development path of immature to mature processes of improved quality and efficiency. The development path is divided into five maturity levels – see Figure 5.14.

The framework shows how IT processes and resources, including application systems, information, infrastructure, and personnel are used to provide the business with infor-

²⁶ www.itgi.org

2. Process Organizations

mation. The framework must have the characteristic of being of value in meeting quality requirements as well as confidentiality, and security requirements. Guidance is provided in mapping individual processes to IT goals and in mapping IT goals to business goals. This provides for the alignment of IT processes with business objectives (van Bon and Verheijen, 2007, p. 148).

COBIT adherence correlates strongly with process improvement because the model has both potent IT controls and potent IT metrics. COBIT provides a single approach to performance in all areas of IT activities. It must be fairly generic to handle each process uniformly. As a result of this, in contrast to ITIL, COBIT does not mandate process flow. ITIL provides more specific information, which can be useful to the performance measurement of ITSM. COBIT is strongly focused on control and less on execution. In summary, COBIT and ITIL are substantially complementary, and one or the other may be more useful to process improvement depending on the area of interest.

Besides effectiveness, efficiency, confidentiality, integrity, availability, and reliability, compliance is not only an explicit criterion, which is crucial in satisfying business objectives. COBIT defines compliance as the process of identifying and following both regulatory document requirements and process standards set by external entities (such as government and trade officials) when planning for the integrated supply chain network (IT Governance Institute, 2007).

ITIL

Consider ITIL, which has two points of interests. It robustly provides details how to perform process-related activities in specific process areas. And ITIL has been widely adopted – and numerous organizations use it (IT Governance Institute, 2008a, p. 36), including two considered in this paper, about which more in due course.

ITIL was originally developed on behalf of the British government by the Central Computer and Telecommunications Agency (CCTA), which was later subsumed by the Office of Government Commerce (OGC)²⁷. ITIL emerged in the mid to late 1980's. The reference model has become a de facto standard in no small measure because it has been continually developed under the influence of the ITSM forum (itSMF)²⁸.

ITIL is a set of guidelines published as a series of five books describing an integrated best practice approach to managing and controlling IT services (Office of Government Commerce, 2007). The Stationery Office (TSO) publishes the material. The latest version ITIL v3 was released in May 2007 (Addy, 2007) and adopts an integrated service life cycle approach rather than being organized around the concepts of IT service support and IT service delivery. Each book is dedicated to one part of the service life cycle.

The book *Service Strategy* (Iqbal and Nieves, 2007) comprises the strategic elements of ITSM (e. g., finance portfolio and service portfolio) and outlines the connection of IT services to business requirements. Design methods of IT services are introduced in the book *Service Design* (Taylor et al., 2007d). *Service Transition* (Taylor et al., 2007c) describes the activities needed to develop the capability to implement new or altered

²⁷ www.ogc.gov.uk

²⁸ www.itsmfi.org

IT services. Duties in the management of service operation are discussed in *Service Operation* (Taylor et al., 2007a). *Continual Service Improvement* (Taylor et al., 2007b) provides techniques for both the adaptation of IT services to ever protean business requirements and the efficiency improvement of IT services. This is why, ITIL is often closely linked with several quality frameworks, such as TQM or Six Sigma. Thus, well defined quality practices might be integrated into ITIL's improvement approach.

Although this life cycle phase offers a generic seven-step improvement approach, it is here posited that it is crucial to provide quality practitioners with more guidance than that which is provided by the seven-step improvement process.

Each life cycle consists of integrated processes. The processes within the life cycle *Service Operation* and *Continual Service Improvement* are described in detail in Section 5.2. Since the other life cycle periods are not substantially dealt with in this paper, interested readers are referred to the relevant books.

Consider the attitude of ITIL toward compliance. A process is compliant in terms of ITIL if the process is implemented as described by the life cycles, and if the process and its results comply with the laws, regulations, and contractual arrangements to which the business process is subject, that is, externally imposed business criteria as well as internal policies. ITIL focuses on efficiency.

2.5. Process Reverse Engineering with Process Mining

This section provides an overview of process mining techniques for the remaining chapters of the dissertation. The overview defines process mining and elucidates its benefits. The overview also introduces both the major personages who have brought process mining and facilities available for the purpose of mining.

2.5.1. Definition of Process Mining

Process mining is a method, which automatically derives the general process knowledge from a completed set of individual instances²⁹, analyzes it from various perspectives, and usually represents it as process models (van Dongen et al., 2005). Hence, it can be assigned to the BPRE approach. Knowledge of underlying business processes reposes in some sort of event log, of which managers or executives may or may not have amply availed themselves. A typical log is recorded by systems ranging from information systems to embedded systems (van der Aalst and Weijters, 2005, p. 236). The event log, also called audit trail or transaction log (Agrawal et al., 1998; zur Muehlen and Rosemann, 2000), stores events occurring during process execution in a structured form. An event indicates a change of state (Kent and Williams, 1999, p. 317), which may be physical (e.g., production of a car) or informational (e.g., arrival of an order). A source of events is provided by such systems or technologies as workflow management systems, ERP systems, CRM systems, RFID events, and web services (van der Aalst and Weijters, 2005, p. 236).

²⁹ In the context of process mining the instance, such as a sales order, is usually known as the case.

2. Process Organizations

During operation, most information systems permit a degree of freedom of executing processes (van der Aalst and Weijters, 2005, p. 236). The scope of freedom is likely to vary with the nature of the underlying process, that is, it ranges from less formalized to highly formalized. An ERP system, for example, allows for varied ways of conducting business transactions. The fact that information systems do not absolutely impose one way of working compels one to ask how employees and applications actually work.

Process mining has many benefits. First, process mining reveals information as to what, how, when, and where something was done – that is, *process discovery*. The aim is to understand what is really happening. Second, process mining can be used for compliance checking, that is, comparing the status quo with what literally ought to be (Rozinat et al., 2009). The way organizations should work can be specified with descriptive processes, notably a reference model. Third, process mining provides guidance in analyzing the performance, a striking example being the detection of a bottleneck. Fourth, process mining can be used for process prediction. Completed instances might be simulated; allowing executives to preview the probable future. Last, all process mining benefits contribute to process improvement. For instance, by comparing the discovered process model with a reference model, discrepancies can be detected and the process improved (van der Aalst and Weijters, 2005, p. 237). An additional contribution to process improvement is the fact that a deep understanding of current processes is considered to be vital for any (re-)design effort (van der Aalst and Weijters, 2005, p. 237). Thus, process discovery might be used as input for a thorough process analysis and for process evaluation during both BPR and CPI activities.

2.5.2. Perspectives

A process can be viewed from different perspectives: process, organization, and instance.

Since the *process* perspective addresses the sequence of activities in which the individual activities are executed, it is also referred to as the control flow perspective. This perspective explores the various possibilities of executing a business process and expresses the choices in terms of process models. It answers questions such as: How are the instances executed? What is the most frequent or the critical way through the business process model? What is the distribution of the instances on the individual ways along the business process? Since the foundation of process mining the process perspective has dominated (van der Aalst and Weijters, 2004).

The *organizational* perspective reconstructs the behavior of the parties by looking at the nature of their participation and the interactions among them. It describes the organizational structure by classifying people in terms of functional aspects (e.g., roles) and organizational units (e.g., groups) and by the relationships among individual performers, including personal relationships. This perspective yields answers for behavior-oriented queries: Which roles exist? What does the communication structure and dependencies look like among the parties? How many employees work on one instance?

The *instance* perspective focuses on the properties of instances. However, this perspective goes beyond the activities being executed or the originators working on an instance. Since instances can also be characterized by more values related to the instance

(e.g., value of a sales order and product type), the instance perspective is intended to establish relations between them. Data mining techniques, which might, for example, monitor correlations between the value of an order and the involvement of specific personnel, is a case in point. This perspective deals with instances, as can be seen from such questions as: To what degree does the instance comply with the business process model? What is the average, the minimal or the maximum throughput times of an instance? How much time is there between the executions of two activities? On the instance perspective in detail, see van der Aalst and Weijters (2005, p. 238).

2.5.3. Representatives

In the middle of the 1990's, the term *process discovery* was coined by Cook and Wolf (1995, 1998). Process discovery analyzed processes to extract automatically models for finite state machines. Agrawal et al. (1998) called the research area *workflow mining* since the mining was used primarily in WFMS to extract workflow models from the execution of workflows. Workflow mining focused on the order in which the individual activities of a process were executed, the so-called control flow. Since 2003 the term *process mining* has been commonly used by many groups (van der Aalst et al., 2003b; Alves de Medeiros et al., 2004). This section both contains a summary of most prominent persons who have analyzed control flow of processes and highlights their approaches. (Alves de Medeiros's dissertation as to process mining approaches may also be helpful – see especially Alves de Medeiros (2006, pp. 15–27).)

Cook and Wolf (1995) dealt with the induction of process schemes for state machines in context of the software development. They first looked at the control flow of sequential processes and expanded their research into concurrent processes (Cook and Wolf, 1998). They used Markov chains and the Ktail approach. With the help of Markov chains they ascertained probabilities for future behavior. The Markov method is predicated on the fact that the probability of the process being in some state depends solely on its previous state (Cook and Wolf, 1995). Therefore the Markov method looks only at the immediate past. For example, as to Markov chain of second order the probabilities for future behavior depend on the past two states the process was in. In contrast, the Ktail approach determines future behavior from historical behavior.

Agrawal et al. (1998) used graph mining to extract workflow models from the execution records of business processes. Their approach assumed that two activities overlapping in time were utterly independent of each other. However, this was simply incorrect – and this premise led to a false conclusion vis-a-vis the predecessor relations among activities. Consider two activities, A and B, with B executed after A, this execution parallel with a third activity, C $[(A \rightarrow B) \parallel C]$. By putting the third activity in an additional predecessor relation that does *not* exist, the relationship is illogical; $A \rightarrow B \rightarrow C$ is wrong. *Golani and Pinter (2003)* extended the work of Agrawal et al. by accounting for the possible dependencies of events overlapping in time – this extension allowed for a more efficient discovery of concurrency.

Van der Aalst and van Dongen (2002) introduced the chronological ordering relations from which the final process model was built. These ordering relations indicate, for

2. Process Organizations

example, whether two activities are executed concurrently or whether an activity is a cause of another activity. The discovered process model is expressed as a Petri net. Since the algorithm does not account for infrequent behavior, it is not robust vis-a-vis noise (Rozinat et al., 2007b). In this regard the mining approach of *Weijters et al.* enlarges the α -algorithm. Its operation resembles that of the α -algorithm, the chief difference being that the ordering relations are inferred based on the frequencies of the instances' occurrences. The algorithms termed α -algorithm and *Heuristics miner* are both implemented in ProM. These mining algorithms are stated in their entirety in Appendix. D.

Ana Karla Alves de Medeiros (2006) used genetic algorithms that emulate evolution. Species presumably arise and adapt by means of natural selection; the analogy to process mining suggests that the algorithm searches for the process model best suited to the instances' behavior. The search begins with an initial population of individuals. The fittest individuals are selected and new individuals are generated by recombining or modifying parts of extant individuals of a given population, or by both of them. The search continues in an iterative process until stop criteria are met. In the context of process mining, an individual represents a process model candidate. The quality of each model is calculated with a fitness function yielding the degree to which the observed behavior is described in the process model. The fitness is the criteria by which the fittest individuals are chosen. Genetic operators (crossover and mutation) recombine these models, creating new potential process models. The approach is robust to noise and grapples with all common structural constructs of process models, among which are duplicates and invisible activities. Furthermore, it both allows for incremental improvement and combinations with other approaches such as heuristics and post-optimization. Note that genetic algorithms require a huge amount of computing time.

2.5.4. Process Mining Facilities

Fueled by ubiquitous event logs, process mining has become a rich research area, and recently, process mining techniques have garnered support from commercial vendors such as Futura Process Intelligence (Futura Reflect, 2009), fluxicon (fluxicon, 2010), Pallas Athena (Pallas Athena, 2010), OpenConnect (OpenConnect, 2009), Iontas (Iontas, 2009), and ARIS Process Performance Manager (ARIS PPM). The vendors' approaches differ in underlying algorithms and in the degree to which process mining is integrated into BPM. Pallas Athena, for example, offers process mining based on genetic algorithms and it is fully integrated in Pallas Athena's BPM suite. In contrast, ARIS PPM's mining technique needs the analyst to model parts of the process at design-time, thus only lending limited support to process discovery.

At this writing, the open source framework ProM³⁰ provides a versatile and expandable environment that can be plugged in for process mining. This architecture makes it easy to add new features without changing the framework itself. The first version of ProM developed by the Eindhoven University of Technology was released in 2004.

³⁰ www.processmining.org

Since its creation the scope of functionalities has become broader, both allowing for the discovery of process models and including functionality, such as process verification, model transformation, model integration, and conformance checking (Verbeek et al., 2006). ProM supports six kinds of plug-ins. *Mining* plug-ins implement mining algorithms with various perspectives. *Import*³¹, *export*, and *conversion* plug-ins interchange data between otherwise incompatible systems – examples of these include – ARIS PPM, different workflow management systems (e.g., Staffware, Websphere), various ERP systems (e.g., SAP, PeopleSoft), simulation tools (e.g., CPN tools) and, notably, analysis tools, such as NetMiner and Viscosity. There are different modeling notations and a wide variety of process models. It is also possible to convert one model to another – for example, a model discovered in terms of a Heuristics net can be mapped to an eEPC (van der Aalst et al., 2007b). After obtaining a process model one of the available *analysis* plug-ins can be used to gauge the suitability of a model, to construct dotted charts, to check the conformance of a model, or to analyze the performance of an instance (Verbeek et al., 2006). *Log filter* plug-ins implement different ways of editing the event log before applying process mining. Concrete examples include the selection of different parts of the event log, the abstraction from infrequent behavior, and the removal of incomplete cases by cleansing the event log.

2.6. Technology Support for Process Organizations

The previous sections have dealt with organizational aspects of managing business processes. Now the focus is on the technology support for process organizations that provides the infrastructure for process improvement applications. Section 2.6.1 starts with an overview of the core technology concepts³². Section 2.6.2 deals with the value of IT. Having emphasized this important aspect of IT, Section 2.6.3 then deals with the IT-Business alignment, addressing as well the suitability of reference models making a contribution to it. Section 2.6.4 deals with compliance and risk. Section 2.6.5 deals with IT governance. RFID technology and its related EPCglobal standard are dealt with in Sections 2.6.6 and 2.6.7 respectively.

2.6.1. Background and Definitions

The concepts information, information technology, information system, information management, IT service management, and IT service warrant classification. Both overviews and detailed explanations of these concepts are given by Addy (2007), Hansen (1996), and Kent and Williams (1999).

Patently, all of these concepts relate to *information*, a term meriting attention. The literature distinguishes among data, information, and knowledge, there is no invariably accepted definition of the term. Nevertheless, this paper uses the one of Teubner (1999,

³¹ ProMimport framework (www.processmining.org) provides further plug-ins for the import of event logs from a multitude of well-known information systems.

³² Note that the development of IT is already placed into a historical retrospective in Section 2.1.1.

2. Process Organizations

p. 17) who refers to information as explicit knowledge that people use or provide. In the computer age, the processing, transfer, and storage of information refer broadly to information technology. It encompasses hardware, data communications, software, and a large variety of input and output devices (Davis and Hamilton, 1993, p. 21). The system, which serves to process and exchange information, is called an information system (Hansen, 1996). Commonly, information systems are configured on the base of a process model, though, crucially, despite the underlying model, most information systems do not completely enforce a specific way of working (van der Aalst and Weijters, 2005, p. 236). The management of information systems and information resources by an individual, a group, or an organization is called information management. According to Davis and Hamilton (1993, p. 22), after the advent of computers, information management has essentially become a new business function with the responsibility to define organizational information requirements, to plan and build an information infrastructure and information system applications, to operate the system, and to organize and to manage these activities.

In contrast to the traditional technology-oriented approaches to IT, a system's technical details are irrelevant to ITSM. Instead, it is a discipline for managing IT operations such as service delivery and service support (Galup et al., 2009). For example, IT software engineering for service operation is outside of ITSM's purview, but the computer systems used for dealing with unplanned interruptions to IT services (i.e., incident management) are in it. A detailed juxtaposition of IT management and ITSM occurs in Section 2.6.5. ITSM deals with the implementation and management of quality IT services to ensure that their delivery meets business objectives in an efficient and effective manner. An IT service is a means of delivering value to customers by facilitating an outcome without their bearing specific costs and risks (Office of Government Commerce, 2007, p. 5). An IT service can be provided to one or more customers by an IT service provider – this IT service is based on the use of IT and supports the customer's business processes (Taylor et al., 2007a).

2.6.2. Value of Information Technology

The subject of IT value (or the business value of IT) has been addressed for many years (Loveman, 1994, p. 45 ff.; Porter and Millar, 1985; Brynjolfsson and Hitt, 1996). A plethora of performance measures had been discussed in understanding the value of IT investments. Examples include: competitive advantage, cost reduction, productivity growth, and profitability increase.

Studies in the 1980's and 1990's predominantly assessed the relationship between IT spending and organizational performance (Loveman, 1994; Brynjolfsson and Hitt, 1996; Barua et al., 1995), but these assessments were generally inconclusive, with results varying from negative to positive returns from IT. Several studies contemporaneous with them evinced the impossibility of drawing any conclusion about IT's effect on organizational performance – a result that became known as the *IT productivity paradox* (Brynjolfsson, 1993). Van Bon et al. (2008, p. 97) neatly illustrate selected organizational performance studies of IT returns.

2.6. Technology Support for Process Organizations

Assessments in inconsistencies of IT value induced a number of researchers (Porter and Millar, 1985; Kohli and Hoadley, 2006; Johannsen and Goeken, 2006; Radhakrishnana et al., 2008) to refine the operationalization of performance variables and to shift their analytical focus to a more process-oriented view. It argues that more convincing evidence of IT value can be derived from investigations of IT impacts on individual business processes or inter-process linkages. Kohli and Hoadley (2006), for example, found empirical evidence that IT investments in tandem with BPR positively and substantially influence performance. And Tallon et al. (2000) proved that the perceived business value of IT increases when IT is more aligned with business strategy. Ravichandran and Lertwongsatien (2005) demonstrated that variation in organizational performance depends on the extent to which IT is used to support and enhance an organization's core competencies. Radhakrishnana et al. (2008) cogently weighed in, seeing IT as the key ingredient distinguishing excellent performance from lesser efforts. The range of process variables they used included inventory levels, capacity use, quality measures, and many others. Porter and Millar (1985, p. 151) opined that IT was extraordinarily significant for the value-added chain and thus also for the activities and processes of an enterprise. Assumedly, the value of IT is also added by supporting business processes and corresponding business process orientation. The IT Governance Institute (2008a, p. 26) stated that stronger IT governance practices correlate positively with better IT outcomes.

The debate received impetus from Carr (2004) who argued that the influence of IT on the success of an enterprise tends to decline. He noted an historical analogy with once-novel advances in communication, production, and logistics. IT packages, like virtually all goods, become accessible to rivals, thereby curtailing the competitive advantages once IT bestowed.

Carr's position continues to be controversially – indeed, recent analyses have reached a conclusion opposite to his. The IT Governance Institute (2009, pp. 12 ff.), for instance, reveals that three-quarters of 250 respondents affirm that IT investments have created value.

A degree of commoditization of IT undeniably exists. Nevertheless, business related concepts of the use of IT can act decisively to contribute value (Johannsen and Goeken, 2006). The support of high-quality and cost-efficient operation of business processes together with the realization of new processes closely aligned with customer demands are the lever for the value contribution. The business related concepts that are important for this dissertation, now to be discussed, are IT-Business alignment, compliance, risk, and IT governance.

2.6.3. IT-Business Alignment

Against the background of the above value of IT the question arises how to apply IT in harmony with business objectives. This challenging question is discussed in the literature under the keyword “IT-Business alignment” (Tallon et al., 2000; Henderson and Venkatraman, 1999). *Yolande E. Chan's* words described it well (Chan, 2002, p. 111):

2. Process Organizations

“The bringing in line of the IT function’s strategy, structure, technology, and processes with those of the business unit so that IT personnel and their business partners are working toward the same goals while using their respective competencies . . . [The bringing in line] is not a state, but a journey.”

This definition is apt – *Chan* captured the process-related character of the alignment and she emphasized that alignment is a perpetual task rather than a state.

Empirical studies showed that improved alignment can lead to higher value contribution. *Chan et al. (1997)* noticed that improvements in alignment increases both the effectiveness and the performance of an organization. *Tallon et al. (2000)* proved that efficient IT management methods result in a significantly higher perceived value of IT for the organization and showed a positive connection between the use of IT and the alignment of IT and business strategy.

Since the end of the 1990’s, a number of alignment models have been developed (*Henderson and Venkatraman, 1999; Chan, 2002; Tallon et al., 2000*) that have tried to explain this phenomenon, especially the Strategic Alignment Model (SAM) from *Henderson and Venkatraman (1999)* (an adapted version is shown in Figure 2.3) has attracted attention.

Henderson and Venkatraman juxtaposed the business side of an organization with the IT. As to both, they distinguished between the strategy (external domain) and the infrastructure (internal domain), and from these two result the four original alignment domains of the model: business strategy, IT strategy, organizational infrastructure and processes, as well as IT infrastructure and processes. The SAM is suitable for the systematization of the alignment problem, but *Henderson and Venkatraman* did not provide any method as to recommendations, techniques, and activities to support the alignment. Control aspects, which generally facilitate the understanding of the dynamic nature of alignment, are also omitted. *Kashanchi and Toland (2006)* have introduced ITIL as an approach to attain IT-Business alignment. In the internal domain of the model, they have replaced IT infrastructure and processes with IT infrastructure and ITIL processes.

In addition to these domains, the SAM model identifies two coordination tasks: strategic fit and functional integration. The former is the vertical coordination of the external domains (business and IT strategy) with the corresponding internal ones. The latter is the horizontal coordination in which the model distinguishes between the strategic (i. e., external domain) and the operative integration (i. e., internal domain).

Alignment in this model means “[...] a balance among the choices made across all four domains” (*Henderson and Venkatraman, 1999*) – their alignment is not merely a bilateral relation but a multilateral one. The alignment starts from the business or the IT strategy as to the neighboring domain. The consideration of the resulting alignment domains business execution, technology transformation, competitive potential, and service level warrant detailed discussion. The impact of ITIL on IT-Business alignment will be determined as well.

Business execution hinges on the alignment among business strategy, organizational infrastructure and processes, as well as IT infrastructure and ITIL processes. The business execution provides the foundation for the design of both the organizational and

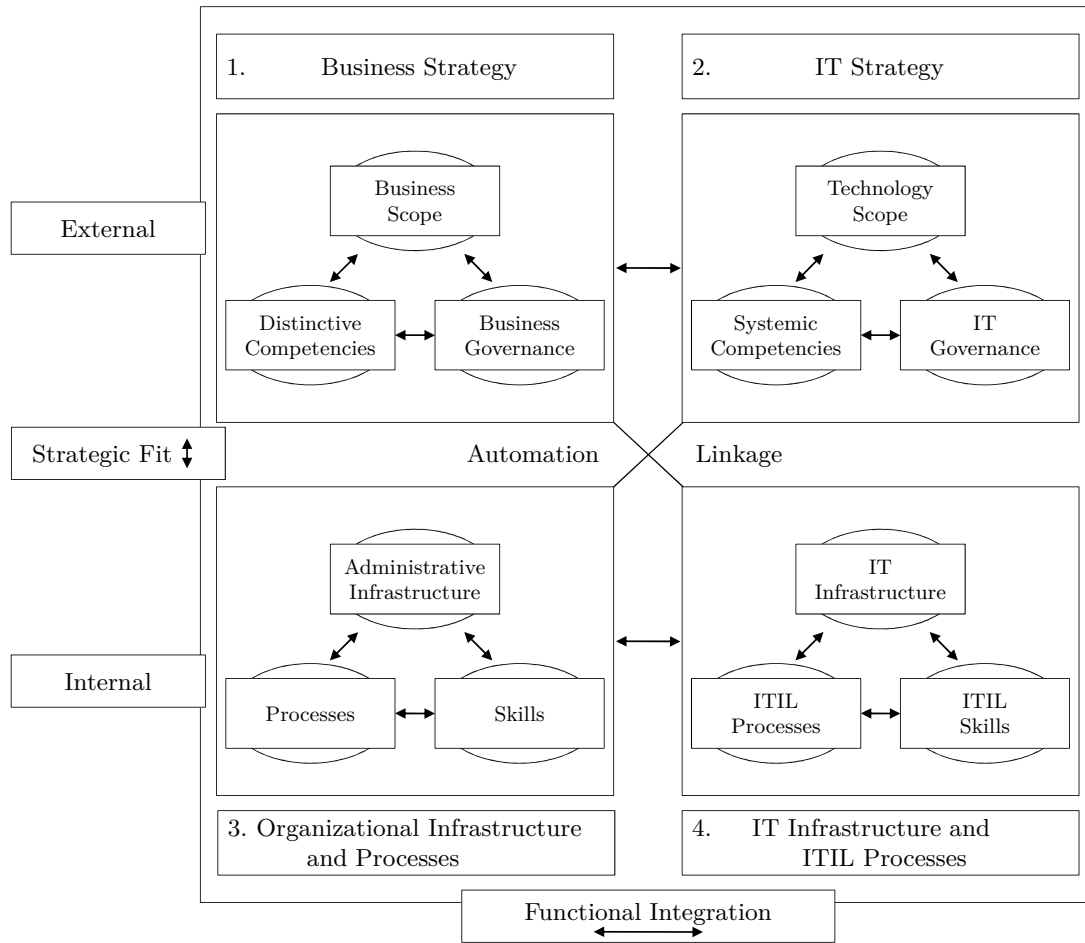


Figure 2.3.: Strategic alignment model according to ITIL (Kashanchi and Toland, 2006)

the IT infrastructure design. Ideally, and presumably in most cases senior executives formulate business strategy – IT specialists design and implement ITIL to enable an organization to transform to a service provider and to ensure the support and delivery of IT services in the best possible manner (Kashanchi and Toland, 2006), albeit with business goals very much in mind. ITIL covers these tasks in the life cycles Service Design and Service Transition.

Technology transformation identifies the alignment among IT strategy, business strategy, and IT infrastructure and ITIL processes. Indeed, ITIL supports the adoption of business strategy through appropriate IT strategy, IT infrastructure, and ITIL processes (Kashanchi and Toland, 2006). ITIL provides much-needed IT vision in the life cycle Service Strategy.

Competitive potential focuses on the alignment among IT strategy, business strategy, and organizational infrastructure and processes. IT supports the implementation of

2. Process Organizations

business strategy through IT capabilities (Kashanchi and Toland, 2006). In the life cycle phase Service Strategy, presumably IT management identifies trends in the IT environment facilitating business managers' understanding of opportunities and threats vis-a-vis IT.

Service level considers the alignment among IT strategy, organizational infrastructure and processes, and IT infrastructure and ITIL processes – the goal is optimal IT service for organizations. As to ITIL, the life cycles Service Operation and Continual Service Improvement certainly aid in enhancing services throughout the organization.

Kashanchi and Toland applied ITIL to the SAM and thereby determined that this reference model affected the four alignment perspectives introduced by the SAM. Their research indicated that ITIL is an effective method for achieving greater IT-Business alignment.

2.6.4. Compliance and Risk

Egregious mismanagements and corporate malfeasance – WorldCom and Enron being stupendously infamous – the economic aftershocks from which the economy continues to suffer have led to a number of new laws and regulations, among which exemplars are Sarbanes-Oxley Act (SOX), Basel II, Gesetz zur Kontrolle und Transparenz im Unternehmensbereich (KonTraG), International Accounting Standard (IAS), and Solvency II. Beeler and Gardner (2006) and Basel Committee on Banking Supervision (BCBS) (2006) give an overview of them. These new mandates typically address various business sectors and aspects, for example, SOX focuses on accountant rectitude, and Basel II seeks to mitigate risks inherent in bank loans. All these mandates strive to ensure transparency, thereby both restoring much lost confidence and increasing the investors' protection.

Naturally, corporate governance and compliance have been in the forefront. The IT Governance Institute (2009) defines corporate governance as a set of responsibilities and practices exercised by the board of directors and senior executives with the goals of providing strategic direction, of ensuring that objectives are achieved, of managing risks appropriately, and of verifying that enterprises' resources are used responsibly. Sackmann and Kähler (2008) state that compliance denotes ensuring that business processes are executed as expected and that both operations and practices are consonant with all laws, regulations, standards – including International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 27000-series, and commercial contracts (e.g., service level agreement, non-disclosure agreement) – and internal governance.

Corporate compliance transcends corporate governance. The latter refers to responsible management behavior. The former refers to every aspect of the organization's behavior.

Where management is responsible for compliance, such as in the case with SOX, Basel II, and Solvency II, activities of governance and compliance overlap. In these cases risk management is especially important. To sustain value, business leaders must consciously deal with risks in terms of corporate governance (Meyer et al., 2003, p. 447).

Since ownership and management are separated or only somewhat overlap in many organizations, danger lies in maliciously opportunistic behavior seeking to maximize profits – so-called “moral hazard” – on the part of management (Schewe, 2005, p. 45 ff.). Risk management cries out for control mechanisms, a notable example of which is the case where shareholders, amply apprised of matters, can attempt to ensure management behavior conforming to their interests.

This congruence of governance and compliance can particularly be observed in the positioning of products of software vendors that are often mislabeled as governance, risk, and compliance (GRC) software. Consider the market for this software; Forrester’s valuations (Rasmussen, 2006, p. 17) indicated that the market volume grew from 85 million U.S. dollars in 2002 virtually sevenfold (i. e., 590 million dollars) in 2006. Stuningly, Forrester forecasted revenue of 1.3 billion dollars in 2011. Examples of software vendors the products of which are claimed to encompass compliance include SAP³³, Oracle³⁴, SAS³⁵, Quadrant³⁶, and Protiviti³⁷.

IT plays an important role in regulations particularly with respect to transparency, safety, and reliability of information. Its role is twofold. IT is instrumental in establishing effective governance and compliance – management information systems and decision support systems, both of which are used by management for decision making, illustrate this. IT itself is also subject to governance. SOX, for example, explicitly addresses control parameters for IT (e. g., SOX Section 404).

To attain compliance, an organization needs to map abstract compliance requirements to concrete control structures and processes, to enforce controls in business operations, and to evaluate effectiveness of controls. Since an organization’s board of directors and senior executives are presumably keenly interested in complying with the new dispensation, ensuring the adequacy and precision of both the system development and the system operation, is more important than ever. Presumably, relentless adherence to compliance guidelines depends on IT governance.

2.6.5. IT Governance

A great many of the new laws and regulations have substantial effects on IT management, as a result of which IT governance has established itself as a counterpart to corporate governance.

Various definitions of IT governance exist. One frequently quoted is that of Weill and Ross (2004, p. 2), the major concerns of which are decision rights and responsibilities. They define IT governance as “the decision rights and accountability framework to encourage desirable behavior in using IT”. In addition to determining who systematically makes and contributes to IT decisions, the *IT Governance Institute* emphasizes procedural mechanisms of IT governance. “IT governance is the responsibility of executives

³³ <http://www.sap.com>

³⁴ <http://www.oracle.com>

³⁵ <http://www.sas.com>

³⁶ <http://www.qrmi.co.uk>

³⁷ <http://www.protiviti.com>

2. Process Organizations

and the board of directors, and consists of the leadership, organizational structures and processes that ensure that the enterprise's IT sustains and extends the organization's strategies and objectives" (IT Governance Institute, 2007, p. 5). Since IT governance describes the part of governance dealing with IT's use, IT governance especially superintends IT management's behavior (IT Governance Institute, 2003, p. 10).

It is certainly acceptable that the ITGI states that following constitutes IT governance: strategic alignment, value delivery, resource management, risk management, and performance measuring (IT Governance Institute, 2007, p. 6). And there is no quarrel with ITGI, who considers compliance to be a major constituent of IT governance (Johannsen and Goeken, 2007).

IT governance contains a spectrum of tasks and therefore IT governance is differentiated from IT management and ITSM (see Figure 2.4). IT management essentially means the effective provision of internal IT services and products, the control of system development and planning, and current IT operation. ITSM itself is centered on the current internal or external customer's needs as to IT's contribution to them. In contrast to IT management and ITSM, the scope of IT governance is broad. IT governance includes present and future demands that arise from both internal and external business sides. External demands are both the customers' and those resulting from laws and regulations (Peterson, 2004).

The plethora of laws with which IT governance needs to comply is a great challenge. Standards and reference models are a great aid in this respect. Two IT governance

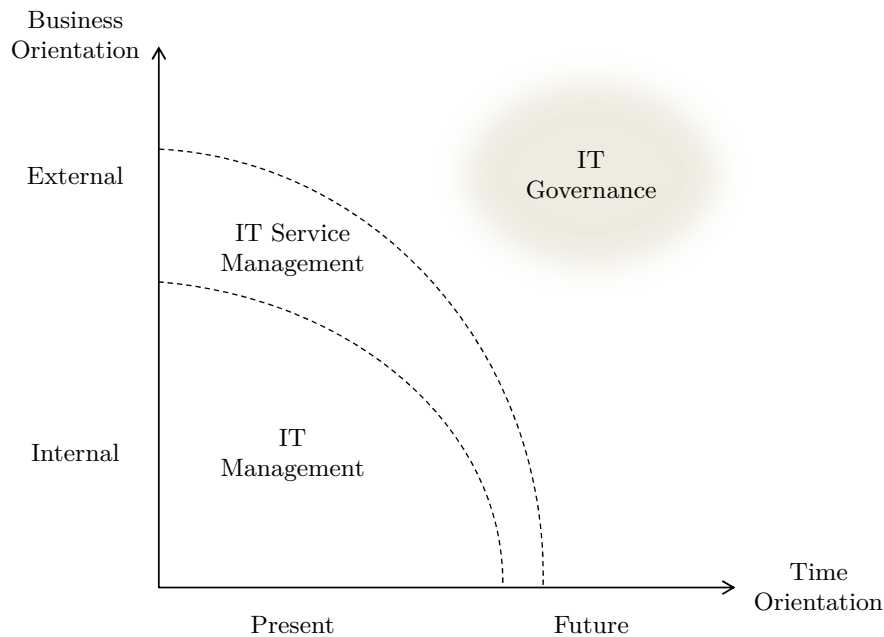


Figure 2.4.: Differentiation between IT management, ITSM, and IT governance (adapted from Peterson (2004))

frameworks, COBIT and ITIL, already introduced in Section 2.4, are frequently employed in industry.

The importance of IT governance is blatantly expressed in recent surveys showing, for example, that 88 % of respondents deem IT governance to be a solution to extant IT problems (IT Governance Institute, 2008b, p. 55). Organizations with above-average IT governance have a 20 % higher return on assets than those with poor IT governance (Weill and Ross, 2004). The correlation between IT governance and IT outcome is also observed by the IT Governance Institute (2008a, p. 19 ff.), which, unfortunately, notes that governance practices are still relatively low.

2.6.6. RFID Technology

Since the reconstruction of product life cycles is herein a research area, and since radio frequency identification (RFID) events represent a possible source that might be exploited for the CPSI approach, this section introduces RFID technology (Tamm and Tribowski, 2010; Günther et al., 2009).

RFID is an automatic identification technology that is used to track locations and movements of objects, especially products. The technology allows for the capture of information, through a large number of sensing technologies, at any time and any place – it has advantages over traditional identification techniques – unlike, say barcode, information can be read contact-less, in bulk, and in non-line-of-sight (Niederman et al., 2007). RFID therefore can speed up information flow and enhance traceability. The technology has been in use for more than half a century, but it has attracted widespread interest and it has been adopted only recently in conjunction with technological improvements. These include reading reliability, storage capability, and costs involved in implementing RFID solutions.

RFID requires different components, namely tags, readers, middleware, and applications. *Tags* (see Figure 2.5) are attached to objects – each stores a unique identifier, the electronic product code (EPC), of the object.

An antenna on the tag emits radio waves generating voltage in the inductor of the passive transponder or triggering the active transponder to send data. The transponder chip, thus activated, sends its EPC to the reader antenna in bit-serial form. *Readers* are

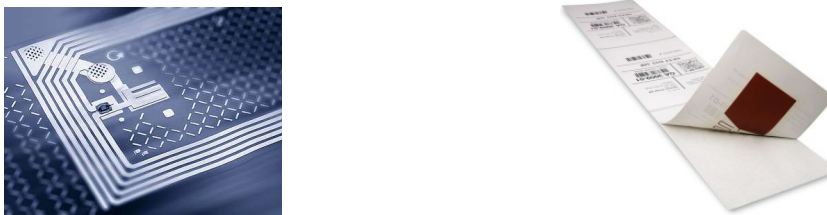


Figure 2.5.: Examples of RFID tags

2. Process Organizations



a) Hand-held reader



b) Gate reader

Figure 2.6.: Examples of RFID readers

hardware devices directly interacting with tags. Hand-held scanners and RFID gates provide everyday examples (see Figure 2.6).

Readers send the data to an extra layer, namely the *middleware* (e.g., SAP Auto ID Infrastructure³⁸ and Microsoft BizTalk³⁹) that collects, filters and aggregates the data, and sends it through the network connectivity to relevant IT *applications*, such as ERP systems, CRM systems, Supplier Relationship Management (SRM) systems, product tracking and tracing systems, and work-in-progress (WIP) tracking systems. Figure 2.7 shows a possible RFID architecture and summarizes concomitant functions.

Currently, international standards are emerging for providing RFID data across the Internet. The most significant is the EPCglobal standard (EPCglobal, Inc., 2007), now addressed in detail.

2.6.7. EPCglobal

EPCglobal is a nonprofit organization founded in 2003 by the standardization consortium Global Standard One (GS1). EPCglobal defines worldwide standards for RFID technology. The *EPCglobal Architecture Framework* describes how the components and standards suggested by EPCglobal fit to form the *EPCglobal Network*. This network is an information network offering guidance to producers, suppliers, shippers, and consumers – information about products is accessed and exchanged. The access to product information is based on EPC, which specifically identifies each product. Various EPC types identify and track different objects; Serialized Global Trade Item Number (SGTIN) identifies trade items, Serial Shipping Container Code (SSCC) identifies dispatch units, and Global Returnable Asset Identifier (GRAI) identifies returnable assets. Uniform Resource Identifier (URI) adheres Figure 2.8's scheme.

³⁸ <http://www.sap.com/solutions/auto-id>

³⁹ <http://www.microsoft.com/biztalk>

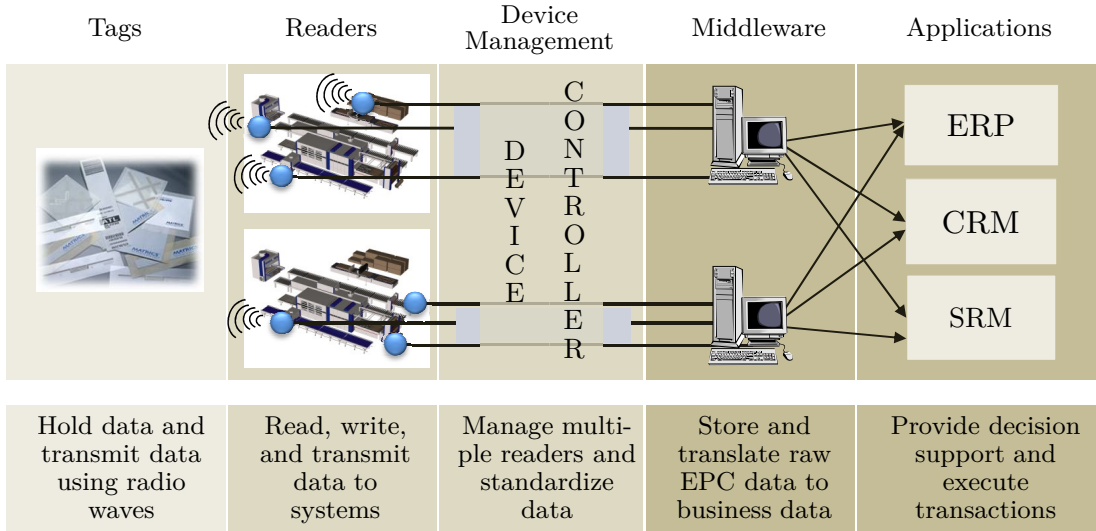


Figure 2.7.: Example of an RFID architecture

The official URI scheme is called Uniform Resource Name (URN) and is registered with the Internet Assigned Numbers Authority (IANA). This name is followed by the terms *epc*, *id*, and the corresponding class of EPC identifier type (i. e., *sgtin*, *sscc*, *grai*). A class of SGTIN EPCs is comprised of the EPC company identifier, the item reference, and the serial number. The company identifier represents a particular company and it is assigned by the GS1. The item reference describes a product class of that company and the serial number refers to a product of this product class. A class of SSCC EPCs is comprised of a company identifier and a serial reference, the latter providing an individual dispatch unit of the company in question. A class of GRAI EPCs is comprised of the company identifier, asset type, and an optional serial number.

An EPC can be attached to products in various ways (e. g., barcode, URN). In contrast to barcodes, the EPC identifies each product instance, not only product categories. EPCglobal also delivers a specification of the electronic product code information services. The EPCIS facilitates the storage and retrieval of information that can be identified through electronic product codes. EPCIS is accessible via the Internet (Günther et al., 2009). The EPCglobal Architecture Framework covers all aspects of reading EPC tag data and of participating in the EPCglobal Network, including the Object Name Service (ONS) and the Product Markup Language (PML). ONS is a global registry of EPC

```
SGTIN urn:epc:id:sgtin:CompanyPrefix.ItemReference.SerialNumber
SSCC  urn:epc:id:sscc:CompanyPrefix.SerialReference
GRAI  urn:epc:id:grai:CompanyPrefix.AssetType.SerialNumber
```

Figure 2.8.: URN notation of the Uniform Resource Identifier

2. Process Organizations

information – PML is language specific to EPC data storage and retrieval (Niederman et al., 2007). Detailed information can be found in EPCglobal, Inc. (2007).

Vis-a-vis electronic product codes reading processes, the term “event” is common. EPCIS standardizes four *event types* (see Figure 2.9). These types capture the aspects of product observations: *ObjectEvent*, *AggregationEvent*, *QuantityEvent*, and *TransactionEvent*.

An *ObjectEvent* is one that occurs when reading one or more electronic product codes. For example, a warehouse entry door reader can always record an event when an item is moved from its entry area. An *AggregationEvent* is used when objects are physically collected or fused. AggregationEvents are created recursively if containers contain other containers. A *QuantityEvent* is one concerned with a known quantity, for example, when inventory is taken. QuantityEvents do not yield EPCs but only an EPC class. A *TransactionEvent* is one in which electronic product codes link with business transactions (e. g., a delivery notification) (EPCglobal, Inc., 2007).

Each event has a number of *fields* – they give precise information as to the current reality of an instant of observation. They are divided into four dimensions: information about the observed object (i. e., EPC, epcList, transactionList, parentID, epcClass, and quantity), the time (i.e., eventTime), the location (i. e., readPoint, bizLocation), and the business context (i. e., process, bizStep, disposition, and action).

The *epcList* is a list of electronic product codes naming the objects to which the event pertained. In case of a TransactionEvent, these objects are linked to the transaction.

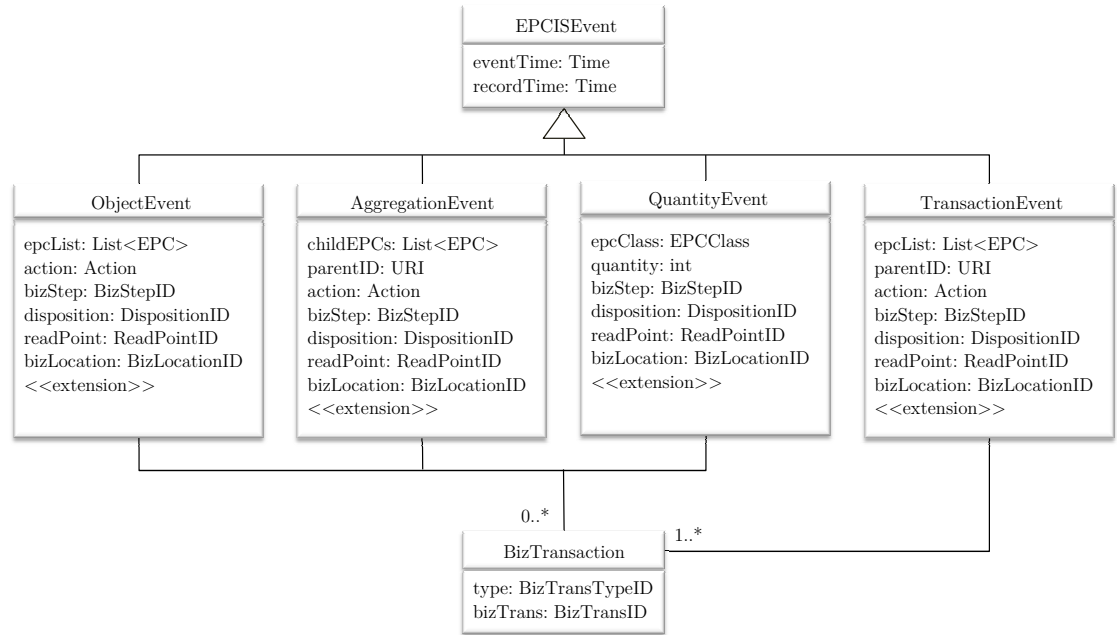


Figure 2.9.: EPCIS events and fields (Tamm and Tribowski, 2010)

2.6. Technology Support for Process Organizations

AggregationEvents and TransactionEvents furthermore specify a *parentID* representing an identifier of the parent of the electronic product codes given in the *epcList*. The event field *type* is an identifier indicating the type of business transactions this business transaction denotes to which the identifier *bizTransaction* refers. An event field refers either to a field defined in the EPCIS specification or to a field defined as an *extension* of this specification. This extension facilitates adaption to a particular situation by adding a new event field to an extant event type. The *eventTime* stores the time at which an event is recorded. The read point at which the event takes place is stored in the event field *readPoint*. The event field *bizLocation* refers to the business location where the objects can be found. The event field *bizStep* reflects the business step of which the event is part. The *disposition* shows the business condition of the objects associated with the electronic product codes. The event field *action* relates the event to the life cycle of the related electronic product codes. The event field can accept the values *Add*, *Observe*, or *Delete*, and captures the semantics of the event. Consider the case of a pallet. When it is created by bundling different boxes, the respective event is an AggregationEvent with action Add. Once it is disassembled, having reached its shipment destination, an AggregationEvent with action Delete is reported. For detailed EPCglobal standard information see EPCglobal, Inc. (2007).

3. Exploiting New Data Sources for Process Mining

As information technology has become ubiquitous, the amount of data has been growing explosively, a phenomenon known as “Big Data”. Organizations today are deluged both with terabytes of data resulting from greater access to data from customers and partners, and with unstructured data flows from a plethora of newly deployed digital devices. Research by McKinsey has calculated that the amount of data has been doubling every eighteen months (Bughin et al., 2010, p. 7). Dave Cappuccio from Gartner (2009) has predicted that the amount of enterprise data will grow about 650 % by 2015.

Fueled by this inundation, process mining has become a rich research area. Since the comprehensive use of process mining is limited to process-oriented data of single-systems, this chapter⁴⁰ investigates how new data sources can be made accessible for process mining as a base for process improvement. The focus is on RFID events and enterprise data as new data carriers.

Section 3.1 reviews related work. Section 3.2 discusses the data preparation necessary to apply process mining. Section 3.3 identifies challenges of mining RFID events and potential means of overcoming them; an algorithm is proposed. Section 3.4 explores the issues organizations are likely to confront with enterprise data, especially that from SAP systems. This section provides solutions and an algorithm to meet the challenges. The new data sources, RFID events and enterprise data, are thereby made applicable to process mining. Section 3.5 concludes and outlines directions of future work.

3.1. Related Work

As this chapter exploits new domains for process mining, this section reviews other approaches, which also target the discovery of processes in a business context, particularly in ERP and RFID environments. This section also assays the literature as to like problems resulting from case construction, about which more in due course.

Turn attention now to current process mining challenges in an ERP environment, particularly those of SAP systems. Research objectives in the area of data preparation stem both from the fact that ERP systems are not process-aware and from the fact that the functional richness of ERP systems makes the underlying data model extensive and complex. These problems are evinced in the works on mapping SAP log data to the Mining Extensible Markup Language (MXML) format. Kassem (2007) developed a method to reconstruct workflow models of an ERP system from trace logs. The trace

⁴⁰ This work is published in Gerke et al. (2009c) and Gerke et al. (2009d).

3. *Exploiting New Data Sources for Process Mining*

files' contents and formats vary from system to system. At this writing, these files do not have standard formats yet, an absence confounding efforts to define general rules for the structure of trace formats. Segers (2007) ascertained that process mining is in fact suitable for extraction of procurement data for auditing purposes, albeit with the proviso that the problems of divergence (i.e., 1:n cardinality) and convergence (i.e., n:1 cardinality) make it more difficult to define a process instance. A single purchase order resulting in a number of goods' receipts provides an example for divergence. Suenbuel and Shan (2006) used a generic framework referred to as the business process exception management to monitor, analyze, and handle exceptions of business processes based on SAP log messages, among which are application logs, job logs, intermediate documents, workflows, and remote function calls. Unfortunately, few SAP applications as to the business process exception management framework exist. Van Giessel (2004) proposed a semi-automated two-step approach in which the relevant database tables are identified first through business objects – the second step, done manually, both retrieves the document flow from the tables and exports the data to MXML. It may well be that the first step lends itself to the work herein. She concluded that data extracting was feasible but that data preparation required arduous effort to locate SAP data sources. Ingvaldsen and Gulla (2007) wrote to similar effect. They described an ERP log analysis system allowing users to define, at a meta level, how events, resources, and their inter-relations are stored and transformed for the use of process mining. Log analysis is related to procurement and logistics.

The above works, sound in themselves, are nevertheless not generally practicable as they result in high manual effort and are only available for specific applications. They may also benefit from elaboration in light of data preparation. No matter whether the data is extracted from tables or transaction logs, the data cannot be directly mapped to activities, which are used in reference models. The different granularities prevent from comparing the extracted processes to the defined ones. This would be desirable with regard to process optimization.

While considerable ground has been broken in intra-organizational analysis, there is only a small body of evidence reporting on inter-organizational case studies. Mărușter et al. (2002) discussed the potential of process mining in supply chains. However, the putative discovery of distributed processes is based on the false assumption that there is a common reference point (e.g., an order number) for all parties. number.

New techniques for data consolidation and analysis are awaited to increase the use of process mining in real-world business settings (Genrich et al., 2007). As to reformatting and enrichment of event log data, Gerke and Tamm (2008) enriched enterprise data by business logic and used clustering techniques to reveal a business process's complexity.

Evidently the literature has no data on the association between EPCIS events and process mining. The correlation relationships given in those events have been reconstructed for this paper, and therefore it is related to the work presented in Wang and Liu (2005), Gonzalez et al. (2006), and Lin et al. (2007). Wang and Liu (2005) presented an integrated RFID data management system, that is, Siemens RFID Middleware, based on an expressive temporal data model for RFID data. The system is predicated on automatically transforming RFID observations into business logic data through user configured

rules. A key concept that containment relationships among objects are used to support tracking and monitoring of RFID objects is very noteworthy. In the similar vein, Lin et al. (2007) explored and took advantage of the containment relationships in relational tables to support special queries in the RFID applications. Interestingly, the model presented by Gonzalez et al. (2006) displayed the design of a highly compressed RFID workflow model that captures typical item movements as well as giving the user the potential to detect, based on duration information, anomalies in them. Noting these can intern be used to detect path segments of unusually long duration, allowing the user either to re-route items or to optimize those segments.

The algorithm set force herein has been devised after observing that certain identifiers, namely electronic product codes, belong to a specific instance. The exploitation of the new data source RFID is unique as it identifies how RFID event data, which is available in the EPCIS format, can be used to construct cases so that process mining can be applied. Furthermore, in amplification of the previously mentioned works, physical flows with logistic bundling operations that cannot be uniquely assigned to one particular physical item have been considered.

This part of the dissertation is comparable to a number of avenues of research that has faced similar problems resulting from case construction. Complex event processing (Luckham, 2001), for instance, has recently received revived attention by those seeking to support process execution in an open and distributed web environment, the challenge being the assignment of incoming messages to the correct instance at runtime. Several works have demonstrated that classical process mining can be applied to web service interaction mining if unique case identifiers are included in each log entry (see Dustdar and Gombotz (2007); Motahari-Nezhad et al. (2007)). As to mining web service interactions, van der Aalst et al. (2008) developed a concept of chained correlation: A message can be assigned to a case via its preceding message, the latter via its own preceding message, and onward in this pattern. If such information is unavailable, cases have to be identified from other indications, though finding these indications is a challenge not only for mining, but also at runtime. In this regard, several executable process modeling languages provide support by offering so-called correlation sets. A correlation set is essentially a query that retrieves identifiers from messages, theses identifiers being unique for a particular instance. The correlation set concept is included in BPEL (OASIS, 2007) and BPMN (Object Management Group, 2006) for dispatching messages to the correct instance at runtime. The correlation problem was discussed in Barros et al. (2007) and Decker and Mendling (2008). There are works concerning mining correlation information including Pauw et al. (2007) that employed an algorithm to identify correlation fields that can be used to construct case identifiers.

3.2. Data Preparation

This section focuses on the data preparation necessary for process mining. It introduces the meta model for process mining data and points to challenges the model poses for data preparation and case construction.

3.2.1. Meta Model for Process Mining Data

The event logs used by process mining algorithms are available in a variety of formats. Most process mining tools share a common XML format usually referred to as MXML⁴¹.

MXML is designed to store recordings of process executions, which typically include information about events referring both to an activity and a case. Information systems also typically record the originator executing the event and, using a stamp, the time at which an event occurred (van der Aalst and Weijters, 2005, p. 236). Take the arrival of an order. This event might contain data related to an arriving order, that is, the customer's name and address, the ordered items and their quantity, the sales agent's name, as well as a time stamp. This information can be stored in MXML because it has a structure⁴² (see Figure 3.1) that accommodates the recordings' content.

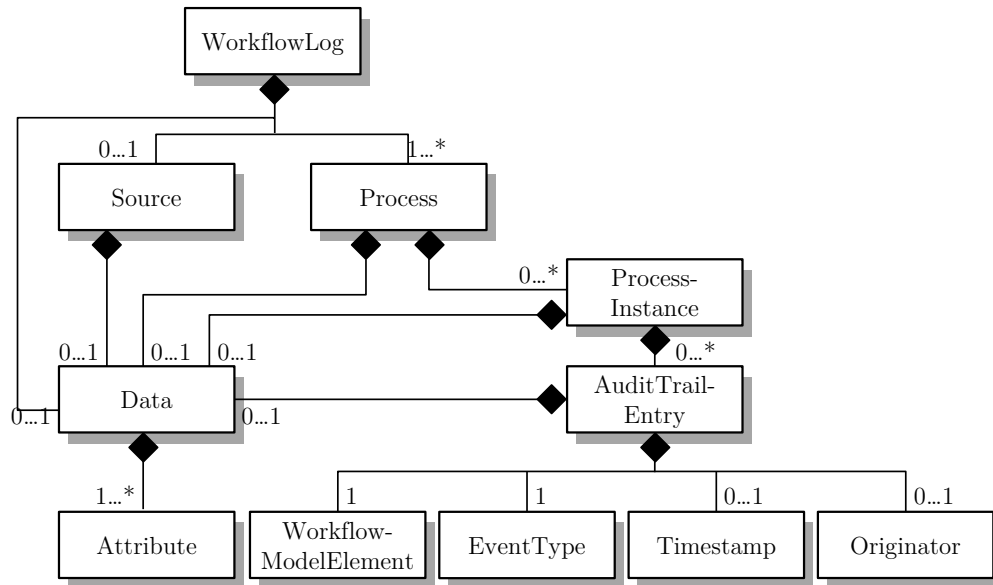


Figure 3.1.: Mining Extensible Markup Language format

The root node of each event log is the *WorkflowLog*, which can contain several *Processes* and optional information about the *Source* as well as further *Data*. Each *ProcessInstance* can have any number of *AuditTrailEntries*. They represent the process events. They ought to be ordered chronologically. The *WorkflowModelElement* describes the process activity. The *EventType* delineates the state change – possible changes being start, assign, resume, and close – of the *WorkflowModelElement*. Each hierarchy level has the optional element *Data*, which can be used to store any additional information. An excerpt of an event log in the MXML format is shown in Section 4.2.4. MXML

⁴¹ Standardizations efforts for a uniform log format have not borne fruit (IEEE Task Force on Process Mining, 2010).

⁴² The schema for the MXML format is accessible at <http://is.tm.tue.nl/research/processmining/WorkflowLog.xsd>.

format is dealt with in greater detail in Verbeek et al. (2006). Van Dongen and van der Aalst (2005) listed the requirements for applying process mining algorithms.

1. Each MXML event should refer to exactly one activity, which should be uniquely identifiable.
2. Each MXML event should happen at a given point in time.
3. Each MXML event should describe the semantics of the activity. (For example, the activity was “started” or “completed”.)
4. Each MXML event should refer to a specific case.
5. Each process instance should belong to a specific process.

The succeeding sections will show that not every data structure of the underlying IS lends itself to dealing with these requirements easily. Consider one involving impediments to *activity identification*. As the activity has the greatest influence on the reconstruction of a process model, one must carefully decide what the activity will be. An activity is chosen or not depending on the different perspectives involved. Requirements two and three lead to a second problem. In process-aware systems each activity is accompanied by a time stamp and an action. Once the activity is identified, these two pieces of information can be mapped to the MXML attributes *Timestamp* and *Event-Type*. If this information is unavailable – and “unavailable” unfortunately includes cases not devoid of information as that term is generally understood, for example, in the case in which a time period but not a point of time is known – the data fails to meet these requirements. This difficulty results in *mapping* problems. Associated with requirement four is the so-called *case construction* problem. In the environment of process-aware systems, the identifier of a process instance is given and changes little throughout the process. It is natural to use this identification (ID) to identify the case. In other environments, however, one seeking the identifier is confounded because single events or activities are recorded without reference to the identifier – this is compounded when a case’s focus shifts throughout the process, making it unclear which case to choose. The last type of difficulty stems from requirement five, that is, *process identification*.

In order to make data available for process mining, four things need to be addressed. First, process mining activities have to be identified. Second, the activities need to be mapped to the MXML format. Third, cases have to be constructed from MXML events. Fourth, cases have to be grouped so that each corresponds to its process. These tasks are crucial because any given strategies may differ substantially from another in its influence on process mining results.

3.2.2. Requirements

The data preparation needs to meet the following requirements with respect to the above activities: activity identification, mapping, case construction, and process identification.

3. Exploiting New Data Sources for Process Mining

- R-1: Correctness.** The conversion needs to be correct. This means that converted MXML events correspond to those of the raw data.
- R-2: Completeness.** The conversion of MXML events needs to be complete in the sense that all essential information of the raw data has to be preserved.
- R-3: Filtering Events.** Filters ideally select only certain activities by specifying parameters. As to EPCIS events, it is possible, for example, to access events from a particular location that are less from the previous 60 days.
- R-4: Efficiency.** Large amounts of data typically have to be processed using process mining algorithms. Therefore, storage space and a reasonable runtime of the conversion algorithm are important considerations.

The following sections analyze the preparation of RFID events and enterprise data.

3.3. Mining Supply Chain Processes

Recently, there have been a number of IT innovations drastically changing the way supply chains are managed. Most notably, RFID technology and complementary concepts, such as the Internet of Things, offer mechanisms for accurate online information sharing, thereby increasing the speed of information flow and creating visibility of information, such as customer demands, inventories, and capacities. Sharing this information helps to overcome problems of the traditional supply chain (Sharma et al., 2007), known as Bullwhip effect (Lee et al., 1997). This effect posits that a small variation in customer demand can lead to a dramatically inordinate increase in demand with which suppliers at the beginning of the chain must deal.

While most organizations overwhelmingly use RFID initiatives to streamline their operations (Ivantysynova, 2008), RFID is promising for much more; it could add value for knowledge creation, decision support, and data mining. By providing precise data as to product location and product characteristics, RFID offers opportunities to analyze supply chain processes (Niederman et al., 2007). Despite the consensus about the need for supply chain analyses, there has been little work on techniques for business processes analyses spanning the supply chain. This may be due to the fact that none of the supply chain parties has a comprehensive overview of the processes being executed in the supply chain.

The value of process mining to supply chains lies mainly in the discovery of causal dependencies of the data at hand, namely data on product locations and product characteristics, as a prerequisite for process verification and improvement. A comprehensive view allows for strategies to discover new relationships and opportunities for process design and redesign (Niederman et al., 2007). Process mining has been successfully applied to various intra-organizational problems (van der Aalst et al., 2007a), but the challenges of mining supply chain processes have not, to date, been richly dealt with. Most process mining algorithms perform well on single-system event logs, which allow tracing individual cases. In many operational environments such case identifiers are not directly

recorded for events. True, RFID and process mining are a promising combination for supply chains – unfortunately, RFID events cannot directly be used for process mining for a variety of reasons.

Section 3.3.1 highlights the challenges of making RFID events available for process mining, in particular case identification and focus shifts are discussed. Section 3.3.2 relates how RFID events conforming to the EPCglobal standard can be used to construct cases so that process mining can be applied, and an algorithm is devised that generates MXML events from EPCIS event logs. In this way, a contribution is made toward applying process mining techniques for supply chain analysis.

3.3.1. Challenges of Mining EPCIS Events

Assuming that every organization in a supply chain uses EPCglobal for the provision of the organizations' RFID data, two problems hinder a direct process evaluation with process mining: case identification and focus shifts.

The problem with *case identification* arises from the fact that electronic product codes are processed. This means that each event relates only to an EPC; therefore there is no explicit case identifier that groups events from the same instance. However, using an EPC as a process ID does not work in every case, as will now be shown.

The use of RFID technology with EPCglobal leads to the *focus shifts* problem. The root cause of this problem is the intertwining of the physical flow of products and the flow of information. An example of focus shifts is seen in the case of the pack and assembly operations depicted in Figure 3.2.

It shows that the focus of a business object varies in time, making it difficult to follow the flow of an instance. The line in the lower part of the figure represents the product

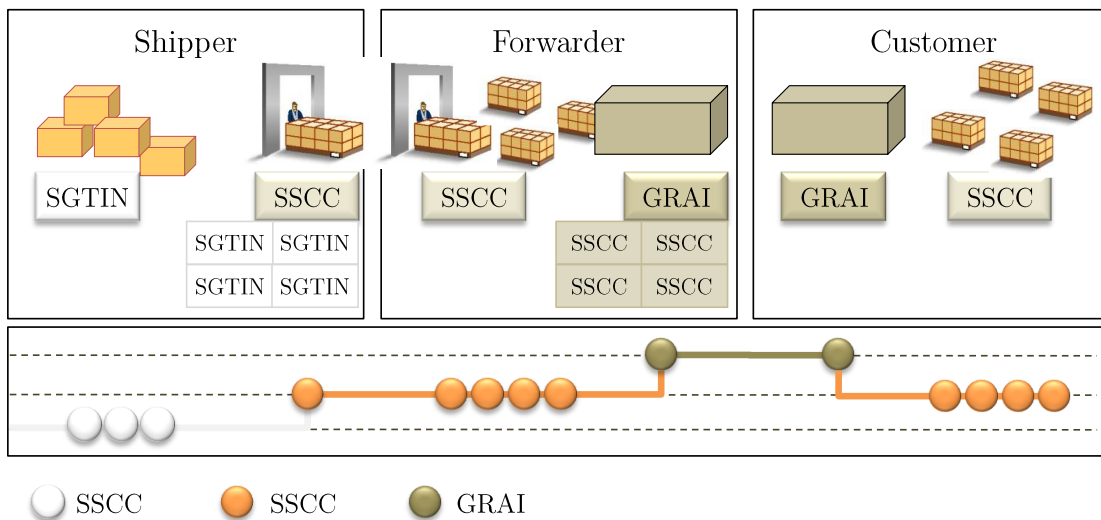


Figure 3.2.: Focus shifts of EPCIS events

3. Exploiting New Data Sources for Process Mining

flow. Whenever the line moves to a higher or lower level, the focus shift causes a gap in tracking the life cycle of the product.

Take the first packaging step at the shipper's site, concluding with the cartons on the pallet. In this process step, an aggregation takes place by which the SGTIN of each carton is associated with the SSCC of the pallet on which the carton is packed. Owing to this aggregation, one can track the exact carton in question though the focus on the business object changes from the single carton to the pallet. Pallets are then loaded into a container identified by a GRAI. The GRAI contains all Serial Shipping Container Codes to identify and verify individual pallets. The container is now the focal point. When it is unloaded, the focus returns to the pallet itself. Monitoring each stage of the flow lets the parties know precisely which products are shipped at a unique pack level. Among other benefits, recall management is facilitated.

Root cause for focus shifts, namely aggregation and transformation will now be dealt with.

Definition 1 (Aggregation)

An *aggregation* is an association of several products. The aggregate typically carries its own EPC. There are two types of aggregation: production and shipment. Production aggregation denotes irreversibility – once different components are truly assembled the end result becomes a new object in its right. In contrast, shipment aggregation is temporary. Its reversal, that is, the breaking up of an aggregate into its components, is called disaggregation. *

Both types of aggregation are transparent and one type can be distinguished from the other based on EPCIS events; production aggregation only includes an aggregation event (*AggregationAdd*), while shipment aggregation has both an aggregation event (*AggregationAdd*) and a disaggregation event (*AggregationDelete*). When an aggregation occurs, its EPC comes into the focus. Once products are disaggregated, the focus returns to the EPC of the individual components.

Definition 2 (Transformation)

There are two types of *transformations*. *Refinement* involves the modification of a primary product into a new one. *Split* involves changing a primary product into two or more final products. The primary product is deemed to be consumed during the transformation. The final products carry a different EPC than the primary product. *

A transformation is not visible vis-a-vis the EPCglobal standard. Although the event *ObjectDelete* (primary product) occurs concurrently with the event *ObjectAdd* (transformed product), the events do not relate to each other. Given the fact that the order in which these two events are captured at a read point is not known, an EPCIS extension field is used to store the EPC of the primary product. This field called *sourceEPC* is provided by the *ObjectAdd* event. The use of the extension of the standard renders moot the question whether this field is inferred from the log or recorded by the EPCIS implementation. The focus is no longer on the EPC of the primary product but to the EPC(s) of the transformed product(s).

There are different ways to deal with these problems. The succeeding section investigates how cases can be constructed from EPCIS events.

3.3.2. Preparation of EPCIS Events for Process Mining

The events serving as input data for process mining are accessible in the EPCIS format described in Section 2.6.7. This section maps these EPCIS events to the MXML format, constructs cases from them, and develops an algorithm to implement these concepts. Specific requirements will now be discussed to justify subsequent verification of the algorithm.

Requirements for Data Preparation

An algorithm has to be developed that retrieves EPCIS events from the EPCIS repository, reconstructs the product flow, and converts the events to the MXML format. A consideration of processing erroneous events, such as faulty electronic product codes, missing electronic product codes, and faulty time stamps is not within the scope of this chapter. In addition to the requirements specified in Section 3.2.2, the following EPCIS-specific requirement needs to be met with respect to the algorithm:

R-5: Product Life Cycle Approach. The preparation of EPCIS events should ideally meet two criteria. First, the resulting MXML structure should allow a process mining algorithm to reconstruct the overall product flow of all EPCIS events. Second, this structure ought to enable the algorithm to extract the life cycle of a certain product class's product.

Mapping EPCIS Events to the MXML Format

EPCIS events' attributes are important for process mining. Information in an EPCIS event can be directly mapped to the MXML event attributes. This is done when the *bizLocation* event attribute translates to the *originator* and the time stamp is stored in the *eventTime* event attribute. Other attributes require detailed consideration that must be taken into account to fulfill the requirements above. The alternatives below are discussed from the point of view of reconstructing the product flow through a supply chain.

Attribute WorkflowModelElement

The MXML event attribute *WorkflowModelElement* represents the activity to which an event is ascribed. This attribute has the greatest influence on the reconstruction of process models by process mining algorithms. An analysis of process mining algorithms has shown that they either focus on the organizational perspective (i. e., use of the originator event attribute only) or on the control flow perspective (i. e., use of the activities event attribute only). Unfortunately, considering a single perspective is insufficient because

3. Exploiting New Data Sources for Process Mining

the representation of the product flows through a supply chain needs to cover information tied both to the location at which events were captured and to the product to which the events belong.

The MXML event attributes candidates that might location related are *bizStep*, *bizLocation*, and *readPoint*.

1. The EPCIS attribute *bizStep* is well-suited to reconstruct workflows because it indicates the business step in which an event was captured. But one cannot tell whether events occurred at different locations or not, much less identify particular sites. See packing activities, which are typically executed by most or all every supply chain parties. If process mining is carried out based on the *bizStep*, information about the different organizations is lost.
2. Granted, EPCIS attribute *bizLocation* is useful to reconstruct in which spatial parts of the supply chain the events took place, but only coarse detail is available, giving rise to information loss. For example, events occurring within one business location are ignored.
3. Presumably there is a natural tendency to equate the reader with a location⁴³ – note that the EPCIS attribute *readPoint* offers the most detailed information about the place at which an event is captured.

Because of the information blurring to which each of the first two candidates is prone, the EPCIS attribute *readPoint* will be considered for the MXML *WorkflowModelElement*. To account for product identification it is necessary to extend the event attribute with product-related information. The level of detail of the product-related information influences the degree to which a process model is generalized. As a result, a three-step granularity has been devised regarding the level of detail of the product classes.

Fine-grained: Information about the product classes is used.

Middle-grained: Information about the EPC product types is used.

Coarse-grained: No product-related information is used.

The different levels of granularity result in different values of the attribute *WorkflowModelElement* as the following example illustrates.

Example 1 (Granularities)

Let an *ObjectEvent* have these attributes:

readPoint = urn:prod:assembly and
epcList = urn:epc:id:sgtin:000001.0000006.457519116749.

The corresponding values of the attribute *WorkflowModelElement* are:

Fine-grained: urn:prod:assembly prodClass:sgtin:000001.0000006

Middle-grained: urn:prod:assembly epcType:sgtin

Coarse-grained: urn:prod:assembly

◇

⁴³ This is based on the EPCIS standard recommendation for the use of read points as physical identifiers instead of logical identifiers (EPCglobal, Inc., 2007, p. 33).

Attribute EventType

Though, as described in Section 2.6.7, the EPCIS standard specifies four event classes, the mapping herein is restricted to `ObjectEvent`, `AggregationEvent`, and `TransactionEvent`. Since `ObjectEvents` and `AggregationEvents` capture information pertaining to one or more physical products, both EPCIS event classes are indispensable for making the data accessible to process mining. `TransactionEvents` provide valuable information to connect the product flows with business aspects. `QuantityEvents` are of subordinate importance. As they do not relate to single electronic product codes, no references to the life cycle of a product can be produced.

Since the MXML format specifies only one event class, *AuditTrailEntry*, this type of event is indicated by the EPCIS attribute *EventType*. The conversion of EPCIS events is based on user-defined MXML event types. These are comprised of the EPCIS event classes and the EPCIS field *action*. Consider the following examples.

Example 2 (Attribute EventType)

Take three events.

An `ObjectEvent` with attribute *action* = *Add* is converted to an MXML event with attribute *EventType* = *ObjectAdd*.

An `AggregationEvent` with attribute *action* = *Observe* is converted to an MXML event with attribute *EventType* = *AggregationObserve*.

A `TransactionEvent` with attribute *action* = *Delete* is converted to an MXML event with attribute *EventType* = *TransactionDelete*. ◇

Although the extant process mining algorithms make no use of the information stored in the *data* tag, this data is especially important for mining algorithms designed for supply chain data. Attributes of the EPCIS events not having any correspondence in the MXML format (e.g., EPCIS expansion attributes), can be mapped to the EPCIS attribute *data*.

The event depicted in Example 3 has been converted at a fine-grained level of granularity and the resulting MXML representation appears in Table 3.1. Table 3.2 shows the overall mapping of event fields.

Example 3 (Mapping)

Let an `AggregationEvent` have the following attributes:

<i>action</i>	=	Add
<i>parentID</i>	=	urn:epc:id:sscc:000001.0000003.333052336329
<i>childEPCs</i>	=	urn:epc:id:sgtin:000001.0000002.839273510945 urn:epc:id:sgtin:000001.0000004.199278144376
<i>bizLocation</i>	=	urn:epc:id:sgln:000001.000004.3
<i>bizStep</i>	=	urn:epcglobal:bizstep:production
<i>readPoint</i>	=	urn:M2.311:Produce:Assembly
<i>bizTransList</i>	=	productionOrderA007000112
<i>eventTime</i>	=	2010-01-20 17:56:12

◇

Table 3.1.: Example MXML event derived from RFID events

```
<AuditTrailEntry>
  <Data>
    <Attribute name="parentID">
      urn:epc:id:sscc:000001.0000003.333052336329</Attribute>
    <Attribute name="childEPCs">
      urn:epc:id:sgtin:000001.0000002.839273510945
      urn:epc:id:sgtin:000001.0000004.199278144376</Attribute>
    <Attribute name="bizLocation">
      urn:epc:id:sgln:000001.000004.3</Attribute>
    <Attribute name="bizStep">
      urn:epcglobal:bizstep:production</Attribute>
    <Attribute name="readPoint">
      urn:M2.312:Produce:Assembly2</Attribute>
    <Attribute name="bizTransList">
      productionOrderA007000112</Attribute>
  </Data>
  <WorkflowModelElement>
    readPoint:urn:M2.311:Produce:Assembly prodClass:sscc:000001
  </WorkflowModelElement>
  <EventType unknowntype="AggregationAdd">unknown
</EventType>
  <Timestamp>2010-01-20T17:56:12.000+00:00</Timestamp>
  <Originator>urn:epc:id:sgln:000001.000004.3</Originator>
</AuditTrailEntry>
```

Deriving Cases from EPCIS Events

Having found a mapping of the EPCIS event attributes to the MXML format, events need to be assigned to instances. Some information systems already augment events with a case identifier during the recording. As the EPCIS events do not have such an identifier, a strategy for grouping instances needs to be developed. The product flow should be reconstructed in such a way that the shifting focus between different assembled and disassembled products of varying granularity is handled appropriately. The following details the strategies and their influence on process mining results.

Strategy 1. An instance contains all events captured at a certain read point.

Advantage: It is possible to reconstruct the parts of a process appearing at that read point.

Disadvantage: The mixing of events referring to different products might distort the mining results so that a causal relationship seems to exist even if only one event occurred per product and read point.

Table 3.2.: Mapping of event fields

EPCIS Event Field	O	A	T	MXML Field
EventClass & action	✓	✓	✓	EventType
EventTime	✓	✓	✓	TimeStamp
ReadPoint	✓	✓	✓	WorkflowModelElement
BizLocation	✓	✓	✓	Originator
BizTransList	✓	✓	✓	Data
Disposition	✓	✓	✓	Data
BizStep	✓	✓	✓	Data
ReadPoint	✓	✓	✓	Data
EpcList, sourceEPC	✓			Data
ParentID, childEPCs		✓		Data
EPCList, parentID			✓	Data

Legend: A = EPCIS eventType AggregationEvent
O = EPCIS eventType ObjectEvent
T = EPCIS eventType TransactionEvent

Strategy 2. Reacting to the first strategy's disadvantage, an instance is made of all events that are captured at a certain read point and are concerned with a certain EPC.

Advantage: This strategy allows for events referring to different products to be distinguished, making it possible to reconstruct the parts of the processes at the individual read points.

Disadvantage: The set of instances fragments into groups of events each of which referring to only one read point. No connections between individual read points can be reproduced – no product flow can be reconstructed.

Strategy 3. An instance includes all events that relate to a certain EPC.

Advantage: The assignment of events to products enables the reconstruction of that part of the product life cycle that is in focus – the product flow becomes partially visible.

Disadvantage: The strategy does not develop relationships with any EPC that relates to the complete life cycle of the product.

Strategy 4. An instance embraces all events that are captured within the life cycle of a product with a certain EPC.

3. Exploiting New Data Sources for Process Mining

Advantage: Connections between the electronic product codes that occur in the life cycle of a product are established. The life cycle of the products can be reconstructed.

Disadvantage: Since an event can be assigned to several instances, redundancy increases the amount of data that the process mining algorithm must process.

Strategy 5. This strategy differs in only one way from strategy four: Here, the instances are divided in different processes, though. A process consists of all instances that cover events of a certain product class.

Advantage: Strategy five has all of strategy four advantages and another. The filters of ProM allow for a targeted reconstruction of the life cycle of products of a certain product class.

Disadvantage: Strategy five has all of strategy four disadvantages.

The first two strategies are undesirable because of the mentioned disadvantages. The expected reconstruction that results when strategy three is used suffers from its drawback vis-a-vis tracking the product. Strategies four and five offer the best for the reconstruction of product flows by a process mining algorithm. Nevertheless, strategy five is adopted and implemented because of its additional benefit above mentioned. The exact assignment of the events to a product life cycle is not trivial. The corresponding algorithm is described in the following section.

Algorithm for Case Construction from EPCIS Events

Herein, a concept similar to so-called *chained correlation* is used. Chained correlations are used in van der Aalst et al. (2008) to assign web service messages to instances. Electronic product codes are used as case identifiers – event entries are added for these items, which are related to a particular EPC. As noted, if a product's EPC is only considered, focus shifts cause gaps in the life cycle of a product. The algorithm closes these gaps by reconstructing the product flow based on all events that are affected by a specific focus shift (strategy five). The rationale behind the algorithm is that events that occur in the context of a product p might also happen in the context of all products that were triggered by aggregation or transformation into the product p . The algorithm called *EPCIS2MXML* can be employed successfully if the following conditions are true:

1. The events are organized in chronological order. Considering the physical relations among products – aggregation and transformation relationships are unrelated – chronological order ensures that the electronic product codes occur in their causal sequence.
2. ObjectEvents with action Add that are generated by a transformation contain the attribute sourceEPC. The attribute stores the EPC of the primary product of the transformation.

Algorithm *EPCIS2MXML* first reads the EPCIS events retrieved from the EPCIS repository via the EPCIS capture interface. While replaying the event log, the algorithm maintains a dependency graph. It manages the relationships among individual electronic product codes. The definition of this graph follows.

Each EPC that appears in an event corresponds to a node of the graph. An aggregation is represented by a directed arc from the parentEPC of the aggregate to all its childEPCs. A transformation is represented by a directed arc from the EPC of the transformed product to its sourceEPC. To each EPC a set of events is attached. Thus, each node and its set of events represent an instance. Since all relationships are of physical nature – all are hierarchically organized. An arc from an EPC epc_1 to an EPC epc_2 can be read as epc_2 *depends* on epc_1 – that is, each event affecting epc_1 also affects epc_2 . This is the case if the product with the EPC epc_2 resulted from the EPC epc_1 by aggregation or transformation. If an event reflects a disaggregation, then the relationship with the corresponding electronic product codes ends after this event and the arc is removed. As an event changes, so do its relationships. The change triggers an update of the whole graph's structure. New nodes are established and arcs may be added or removed from the graph. The events are propagated through all descendants in the graph, meaning that any event is not only attached to all electronic product codes occurring in the *epcList* or *parentID* field of the event but also to all electronic product codes that *depend* on them. The latter attachment ensures that an AggregationEvent has affected both the aggregate and the primary products. Each event that affects a transformed product is also bound to its source product.

Because many electronic product codes can occur autonomously in an event log, the resulting structure of the dependency graph is a forest.

The algorithm *EPCIS2MXML* is defined in pseudo-code. The algorithm's salient aspects merit emphasis. Line 3 captures the transformation relations. If the event generator is able to handle the EPCIS extension field *sourceEPC*, naturally the algorithm works; if it is unable, the algorithm works albeit with transformation steps unidentified. The function **addEvent** assigns every event e , which appears in connection with an EPC, to the node epc (e.g., line 9). Simultaneously, it is propagated through the corresponding tree so that e is also assigned to all nodes in the partial tree, which has its root in the node epc . The propagation takes place in the function **propagateToDescendants** (e.g., line 10). Lines 17 to 19 account for the reversal of an aggregation relation. After assigning the event to all descendants, the interdependence of *parentEPC* and *childEPC* is eliminated. The need for the update of the graph's structure stems from both functions **addChildren** and **removeChildren** in lines 15 and 19.

Once the events are mapped and grouped into cases according to the assignment strategy, the last step of the conversion of EPCIS events can be taken – the MXML structure is written from the dependency graph. For each node in the graph an instance in the MXML tree is created and all attached events are enclosed as child nodes.

Grouping instances in distinct processes identified by EPC classes allows mining for all products on their way through the supply chain as well as for specific product classes.

Algorithm *EPCIS2MXML* is implemented as a plug-in for ProMimport. The plug-in queries events from an EPCIS repository and generates MXML events according to the

3. Exploiting New Data Sources for Process Mining

EPCIS2MXML Case construction of EPCIS events

INPUT: events – chronological ordered EPCIS events

OUTPUT: cases – map from eEPCs to sets of events

```
1: for all Event e in events do
2:   if e.isObjectEvent then
3:     if e.isAddObject with sourceEPC then
4:       for all EPC epc in e.EPCList do
5:         graph.addChild(epc, sourceEPC)
6:       end for
7:     end if
8:     for all EPC epc in e.EPCList do
9:       graph.addEvent(epc, e)
10:      graph.propagateToDescendants(epc, e)
11:    end for
12:  else if e.isAggregationEvent then
13:    graph.addEvent(e.ParentEPC, e)
14:    if e.isAddAggregation then
15:      graph.addChildren(e.ParentEPC, e.ChildEPCs)
16:      graph.propagateToDescendants(e.ParentEPC, e)
17:    else if e.isDeleteAggregation then
18:      graph.propagateToDescendants(e.ParentEPC, e)
19:      graph.removeChildren(e.ParentEPC, e.ChildEPCs)
20:    end if
21:  else if e.isTransactionEvent then
22:    graph.addEvent(e.ParentEPC, e)
23:    graph.propagateToDescendants(e.ParentEPC, e)
24:    for all EPC epc in EPCList do
25:      graph.addEvent(epc, e)
26:      graph.propagateToDescendants(epc, e)
27:    end for
28:  end if
29: end for
30: cases = graph.nodes
```

algorithm. The query is controlled by a number of parameters, such as time intervals, electronic product codes, locations, and event types. Excluding TransactionEvents, for example, derives a process model that purely reflects the product flow. The resulting MXML file is then be provided to ProM as a starting point for process mining algorithms.

Algorithm *EPCIS2MXML* closes the gap in the life cycle of products, thereby offering optimal prerequisites for the reconstruction of process models.

Note that the algorithm does have a disadvantage – by propagating events through the dependency graph, redundancy is added. Events are often copied, sometimes repeatedly, depending on the complexity of the supply chain (e. g., the length of the paths and the

branching degree by aggregations). Events are also duplicated if they appear in the context of several EPCs.

3.4. Mining Enterprise Processes

Given the fact that enterprise systems enable the execution of business processes, these systems' data provide insight (Ingvaldsen and Gulla, 2007). Any ERP system, for instance, often is the backbone operation of an organization. According to the definition provided by Wallace and Kremzar (2001), an ERP system is

“an enterprise-wide set of management tools that balance demand and supply [...] containing the ability to link customers and suppliers into a complete supply chain, employing proven business process for decision-making, and providing high degrees of cross-functional integration among sales, marketing, manufacturing, operations, logistics, purchasing, finance, new product development, and human resources, thereby [...] enabling people to run their business with high levels of customer service and productivity, and simultaneously lower costs and inventories.”

Among the major suppliers of ERP systems are SAP, Oracle, and Microsoft. SAP will be examined – it has emerged globally as a major product in the market for enterprise systems (Jacobson et al., 2007). The SAP ERP system is augmented with additional systems, such as CRM, Supply Chain Management (SCM), product life cycle management (PLM), and SRM, forming the SAP Business Suite.

Section 3.4.1 addresses major challenges for applying process mining algorithms to enterprise data: the complexity of the database, problems related to the concept of transactions, and focus shifts caused by associations. Section 3.4.2 shows, using the example of complaint management, how enterprise data can be used to construct cases. An algorithm is developed the outcomes of which are MXML process mining events from data stored in a CRM system. A contribution is made toward applying process mining techniques for enterprise analysis.

3.4.1. Challenges of Mining Enterprise Processes

Enterprise systems long have been exploited merely as an information source for process mining. This stark limitation has had its reasons.

First, the complexity and size of SAP tables in an enterprise-wide relational database make it extremely difficult to determine how all the data and dependencies are stored. Not only do numerous tables have to be joined – the data relevant for process mining has to be located in each table (Ingvaldsen and Gulla, 2007). Consider a CRM activity, that is, an activity that is undertaken by staff members on behalf of an organization. It stores a variety of information resulting from interactions between them and customers during the relationship life cycle. These CRM activities include telephone calls, letters, customer visits and contacts, and preliminary tasks. Example 4 shows how a business partner involved in a CRM activity can be selected.

3. Exploiting New Data Sources for Process Mining

Example 4 (Table Selection)

Let a CRM activity be identified by its unique Globally Unique Identifier (GUID) E2ADEED3284ED44B8162DD776A2A67A8.

Table	Key
crmd_orderadm_h	E2ADEED3284ED44B8162DD776A2A67A8 = crmd_orderadm_h-guid
crmd_link	crmd_orderadm_h-guid = crmd_link-guid_hi
crmd_partner	crmd_link-guid_set = crmd_partner-guid
but000	crmd_partner-partner_no = but000-partner_guid
crmc_partner_ft	crmd_partner-partner_fct = crmc_partner_ft-partner_fct

◇

The first table *crmd_orderadm_h* provides information related to the CRM activity that is specified by the *GUID*. This identifier then serves as input to retrieve the field *guid_set* from the table *crmd_link* containing the link between the activity and the business partners. The field *guid_set* in turn provides the means of accessing the fields *partner_no* and *partner_fct* from the table *crmd_partner*. The *partner name* is selected by the former field from table *but000* – the latter field provides the *partner function* (e. g., responsible agent and group) from table *crmc_partner_ft*. The name of the business partner and the description of the partner function are ontological information that aids in interpreting otherwise indecipherable values such as the GUID and the partner number. This elementary example demonstrates that a grand total of five tables have to be joined to extract business partners if the data base is accessed conventionally. Indeed, collating the data is even more time-consuming and difficult when the process starts at inquiry and runs the whole gamut through quotation, ordering, shipping, and final billing of products.

A further difficulty also stemming from complexity is that the SAP Business Suite consists of highly integrated software packages. Each performs specialized business functions and typically stores each piece of data only once, meaning that a process might be split into a number of applications of the SAP Business Suite. For example, customer contacts and marketing activities are processed in the CRM system and the corresponding sales orders are handled in the ERP system. Therefore, it is necessary in certain cases to collate the data from more than one enterprise system.

Second, database transactions⁴⁴ typically used in an enterprise system are problematic vis-a-vis activity identification and case construction. On top of the tables of the SAP enterprise system, most of the business processes operate on documents with which the user can modify the tables of the underlying database. To avoid discrepancies when two

⁴⁴ A transaction advances a system from one consistent state to another. In conformance with the ACID principle (atomicity, consistency, isolation, and durability), all data changes between the start and the end of a transaction are processed as an atomic unit.

or more database tables are updated, the updates need to be stored in the database simultaneously. SAP ensures data consistency by using the concept of transactions. A transaction either commits or aborts – all actions of the transaction are either performed and stored or they are undone.

This “black and white” model, success or failure, is confounding because numerous operations are executed as a whole. See Figure 3.3’s example – the document *Sales Order* provides an understanding of how the example is processed in the enterprise system.

This document consists of a *header* and a tree of associated *items* each of which embraces a set of attributes. The header attributes, such as ID and posting date, are associated with all item entries. Product, quantity, and amount are examples of item attributes. The items refer to a tree of additional attributes, such as schedule lines. The semantic relationship between the header and its items is defined by a unidirectional parent-child association.

Definition 3 (Composition)

Associations that link the different items of a single document are *compositions*. A compositional association represents a strong semantic relationship between parent and children within one document, meaning that one or more instances of the children depend on the existence of one instance of the parent. One instance of a parent may have zero, one, or more than one associated instances of the corresponding children. *

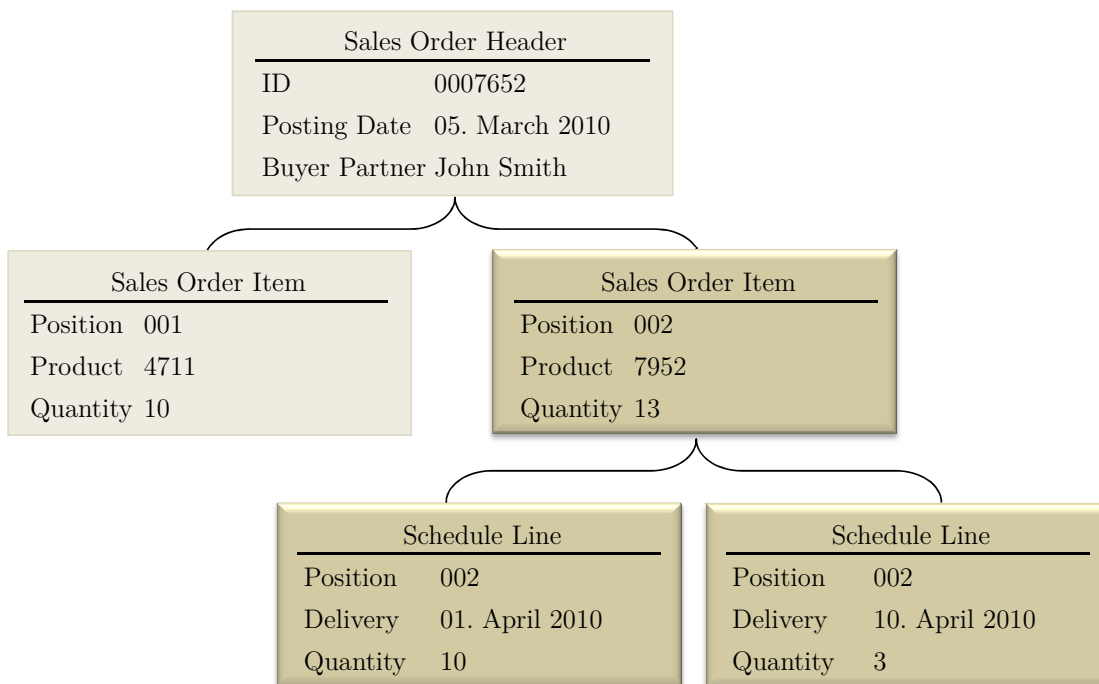


Figure 3.3.: Composition within an business object

3. Exploiting New Data Sources for Process Mining

Example 5 (Composition)

Take the SCOR activity D2.2 *Receive & Validate Order*. It is comprised of the creation of the sales order, the reservation of required inventories, and the scheduling of the delivery date. The document's header *Sales Order* has at least one compositional association to the *sales order item*, which refers to the reserved products. The sales order item itself has an association to one or more *schedule lines* and their concomitant delivery dates.◇

Definition 3, like Definition 1, is concerned with aggregation. Both composition and aggregation are associations that can be characterized by a whole-part relationship between the objects involved. A composition is a strong form of an aggregation because a part can be assigned to one whole at one time. The cardinality cannot be greater than one. Naturally, as to RFID events, the cardinality can be greater than one.

Given the fact that the parent, that is, the document itself, is identifiable but the semantical relationship with its children (e.g., items and schedule lines) is not, direct business process analysis with current process mining tools is prevented. Example 5 showed that a transaction normally incorporates a series of user interactions and therefore might span a number of process mining relevant activities. This example appears to indicate that it is essential to seek for activities not only on the parent level (i.e., sales order creation) but also on the corresponding children level (e.g., scheduling of the delivery date).

The atomicity principle is not only challenging for activity identification but also for identifying information about the point of time at which one specific user interaction was performed. Since precise time stamps are important pieces of information from which the process mining algorithm derives assumptions about ordering of activities, it is hard to meet the requirement specified in Section 3.2.1.

In addition to compositional associations, there are associations across different documents. Take the processing of a sales order. It is created in the IS, preferably with reference to a quotation or an inquiry maintained earlier in the system, or to both of them. After ordered products are manufactured they are shipped with a link to the sales order. Finally, the goods issue refers to the delivery. Example 6 below shows that not every document is an individual document – this document is created as part of a series of related documents. Data flow from one document to another contributes to the cross-functional integration among the key functions including sales, production, logistics, purchasing, and finance. Data flow reduces manual activity. Note: data flow is very waxing vis-a-vis case construction because each new document represents a focus shift, it shifting each time a document is created.

Definition 4 (Cross-Document Association)

An association that links the relationship between two nodes of different documents is called *cross-document association*. *

The relationship between successive documents of a business process is stored in the enterprise system as a link entry in the document flow. It is therefore possible both to see in which order the documents within the business process were created and to see the

preceding and following documents. The relationship, which can occur on both header and item levels, is usually not uniquely identified.

Example 6 (Cross-Document Association)

During order processing, it is sometimes necessary to split an order into two or more deliveries. Delivery splits can occur on order item level due to, among other things different ship-to-parties, delivery dates (see Figure 3.3), and routes. A delivery split might also become necessary during delivery processing because of, for example, exceeding loading capacity limitations. ◇

3.4.2. Preparation of Enterprise Data for Process Mining

This section investigates how data reposing in an SAP CRM system can be prepared for process mining for enhanced supply chain analysis. Preparing this data for process mining requires mapping of data to the MXML format (van Dongen and van der Aalst, 2005) and a construction of cases. An algorithm is introduced as a vehicle to realize the concept. Remarks regarding validation of the algorithm's implementation conclude.

It is necessary to provide a description of the data handling in the SAP CRM system. Data handling in an SAP CRM system is similar to that of an SAP ERP system – both create documents within an application. By introducing a database abstraction layer, which is labeled the business object layer (BOL), SAP restricts the available database functionalities. In the BOL, business processes operate on business objects (BOs). A BO, such as a sales order and a business partner, is a representation of a type – not an instance level – of a uniquely identifiable business entity. A business process is provided by one or more business objects. Their business data is described as a set of attributes in the business object repository (BOR), known as BOR objects. BOR is the object-oriented repository of R/3 data and processes embracing the business object and their components, such as methods (known as Business Application Programming Interface (BAPI)), attributes, and events. A BAPI accesses the application functions. Business logic and business data is encapsulated. Only the interface functionality through which BAPI is exclusively accessible is visible to the user (Füchsle and Zierke, 2009). Naturally, BAPI has become standard – SAP guarantees its stability regarding content and interface. After a BAPI is released by SAP, the interface definition and parameters stay the same, ensuring that the application program is not influenced by changes of the underlying R/3 software and data. And BAPI is accessible for all applications and information systems that support the SAP protocol remote function call (RFC). The BO layer is inherent in many packages, such as CRM, of the SAP Business Suite. Fortunately, BOs are already available in the SAP ERP because many business entities employ them.

As a result of this change in data handling, independent entities, such as business partners and products, are BOs in their own right. To address this, relations between different business objects can be established by extending cross-document associations.

Definition 5 (Cross-BO Association)

An association that links the relationship between two nodes of two different business objects is called *cross-BO association*. *

3. Exploiting New Data Sources for Process Mining

In Figure 3.4, the header node of BO *Sales Order* has a cross-BO association to the header node of the BO *Business Partner* for linking customer data to the sales order header, and the node sales order item has a cross-BO association to the header of BO *Product*.

The next section specifies requirements the conversion of enterprise data to MXML events should satisfy.

Requirements for Data Preparation

As described in Section 3.2.2, the preparation of the enterprise data needs to be correct, complete, and efficient. Filtering parameters, such as time intervals and activity types, must provide partial access to specific data. And the following enterprise data-specific requirement must be satisfied with respect to the converting algorithm.

R-5: Representation of Reference Activities. As the enterprise system has no attributes providing activities the granularity of which is comparable to that of a reference model, they need to be inferred from CRM transactions.

Identifying Activities from Enterprise Data

One gleans much from reference models and one so appraised ought to be convinced that there are important activities that should be represented in the process model though there is no attribute currently doing so. The need to identify activities from enterprise data stems from the transaction concept. Because a business transaction encapsulates a unit of activities, they must be derived from one or more attributes available as to that transaction. The single fine-grained activities obtained should be juxtaposed with those of a reference model, preferably that of ITIL. This reference model is ideal because the business scenarios of the CRM application (about which more

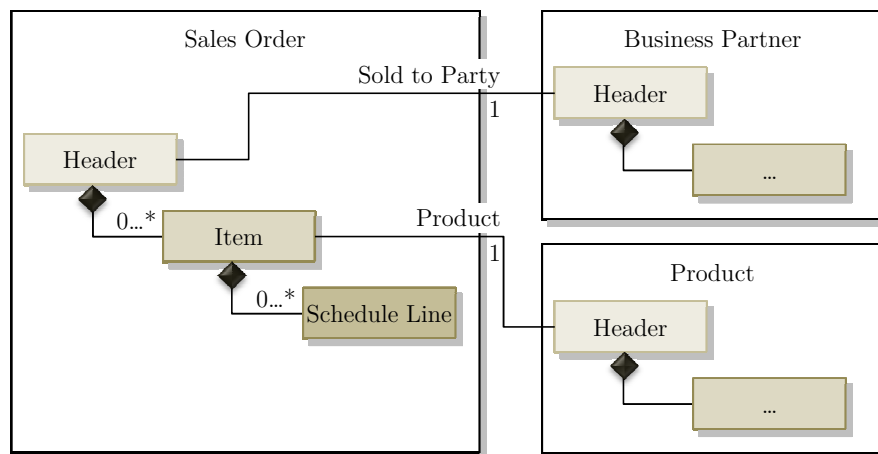


Figure 3.4.: Associations among business objects

in Section 4.1.3) substantially correspond to the recommendations for operations of a service desk proposed by ITIL (Redinger, 2007). The Cross Industry Standard Process for Data Mining (CRISP-DM) recommends deriving activities, especially if certain facts are not being covered (Cross Industry Standard Process for Data Mining, 2008, p. 50).

The construction of fine-grained activities will be described in more detail. The main business transaction is the CRM activity. In addition to the attributes of this business transaction, also employed are documents attached to the CRM activity, ontological information, and change documents to extract the activities. Note the difference between a CRM activity (i. e., one business transaction) and an activity extracted from it (i. e., an activity as defined in Chapter 2.1.2). The construction of derived activities is based on the concept of inheritance, which is used in the context of object-oriented modeling. If an attribute of the business transaction itself is not associated with properties, such as the time stamp and the originator, it inherits those of the business transaction to which the attribute belongs. An example follows.

Example 7 (Identifying Activities)

Let a complaint have the following attributes stored in one business transaction, that is, one instance of the BO CRM activity:

```

type          = Z003 (Customer Relations)
reason        = Suitcase lost
created at    = 20.01.2009 13:24:44
created by    = IC Desk
changed at    = 22.01.2009 16:33:35
changed by    = IC Desk
status        = E0007 (Customer request), 22.01.2009 16:33:35,
               IC Desk
status        = E0005 (Closed), 23.01.2009 09:13:35, IC Desk

```

In addition, an outgoing e-mail messaging is attached to the CRM activity. The attachment is associated with the time stamp 22.01.2009 16:12:25. ◇

The fine-grained activities extracted from the complaint depicted in Example 7 are listed in Table 3.3, and merit discussion. Activity *type Z003* is first enriched with ontological information from table `crmc_proc_type_t` (text for transaction types) to reveal the CRM activity type *Customer Relations*. This field in combination with the fields *created at*, *created by*, and *status* is translated into the first activity *Create activity Cust. Relations* and its associated attributes time stamp, originator, and status. The second activity *Classify problem* is triggered by filling out field *reason*. The time stamp, the originator, and the status are adopted from the initial creation of the CRM activity. The attachment to the CRM activity results in the third activity. Because the e-mail is outgoing, the activity *Write mail*, with the associated time stamp, originator, and status, is produced. The change of the status to *Customer request* updates the fields *changed at*, *changed by*, and *status*, as a result of which the activity *Edit activity Cust. Relations* is concatenated with the fields *type*, *changed at*, *changed by*, and *status*. Every

3. Exploiting New Data Sources for Process Mining

Table 3.3.: Activities extracted from a CRM activity

Activity Description	Time Stamp	Originator	Status
Create activity Cust. Relations	20.01.2009 13:24:44	IC Desk	complete
Classify problem	20.01.2009 13:24:44	IC Desk	complete
Write mail	22.01.2009 16:12:25	IC Desk	complete
Edit activity Cust. Relations	22.01.2009 16:33:35	IC Desk	complete
Ask customer	22.01.2009 16:33:35	IC Desk	complete
Close contact	23.01.2009 09:13:35	IC Desk	complete

change in status is stored in the change documents providing the status, time stamp, and originator. Table tj30t provides the text of the status. From these change operations stem the activities *Ask customer* (i.e., status E0007) and *Close contact* (i.e., status E0005).

Mapping Enterprise Data to the MXML Format

All attributes can be directly mapped from Table 3.3 to the MXML format. The activity description corresponds to the event attribute *WorkflowModelElement*, the originator translates to the event attribute *Originator*, the time stamp is stored in the *EventTime* event attribute, and the status in the event attribute *EventType*. Instances are gathered in the process *Interaction Center (IC)*.

The MXML representation of the converted event depicted in Example 7 is shown Table 3.4.

Table 3.4.: Example MXML events derived from enterprise data

```
<AuditTrailEntry>
  <Data>
    <Attribute id="activityID">137</Attribute>
    <Attribute guid="GUID">
      E2ADEED3284ED44B8162DD776A2A67A8</Attribute>
    </Data>
    <WorkflowModelElement>Create activity Cust. Relations
    </WorkflowModelElement>
    <EventType unknowntype="complete">complete</EventType>
    <Timestamp>2009-01-20T13:24:44.000+00:00</Timestamp>
    <Originator>IC Desk</Originator>
  </AuditTrailEntry>
</AuditTrailEntry>
```

to be continued on the next page ...

Example MXML events derived from enterprise data – continuation

```

<Data>
  <Attribute id="classification">Lost luggage</Attribute>
  <Attribute guid="GUID">
    E2ADEED3284ED44B8162DD776A2A67A8</Attribute>
  </Data>
<WorkflowModelElement>Classify problem</WorkflowModelElement>
<EventType unknowntype="complete">complete</EventType>
<Timestamp>2009-01-20T13:24:44.000+00:00</Timestamp>
<Originator>IC Desk</Originator>
</AuditTrailEntry>
<AuditTrailEntry>
  <Data>
    <Attribute guid="GUID">
      E2ADEED3284ED44B8162DD776A2A67A8</Attribute>
    </Data>
    <WorkflowModelElement>Write mail</WorkflowModelElement>
    <EventType unknowntype="complete">complete</EventType>
    <Timestamp>2009-01-22T16:12:25.000+00:00</Timestamp>
    <Originator>IC Desk</Originator>
  </AuditTrailEntry>
<AuditTrailEntry>
  <Data>
    <Attribute id="activityID">137</Attribute>
    <Attribute guid="GUID">
      E2ADEED3284ED44B8162DD776A2A67A8</Attribute>
    </Data>
    <WorkflowModelElement>Edit activity Cust. Relations
  </WorkflowModelElement>
    <EventType unknowntype="complete">complete</EventType>
    <Timestamp>2009-01-22T16:33:35.000+00:00</Timestamp>
    <Originator>IC Desk</Originator>
  </AuditTrailEntry>
<AuditTrailEntry>
  <Data>
    <Attribute guid="GUID">
      E2ADEED3284ED44B8162DD776A2A67A8</Attribute>

```

to be continued on the next page ...

3. Exploiting New Data Sources for Process Mining

```
Example MXML events derived from enterprise data – continuation
</Data>
<WorkflowModelElement>Ask customer</WorkflowModelElement>
<EventType unknowntype="complete">complete</EventType>
<Timestamp>2009-01-22T16:33:35.000+00:00</Timestamp>
<Originator>IC Desk</Originator>
</AuditTrailEntry>
<AuditTrailEntry>
  <Data>
    <Attribute instId="instanceID">36</Attribute>
  </Data>
  <WorkflowModelElement>Close contact</WorkflowModelElement>
  <EventType unknowntype="complete">complete</EventType>
  <Timestamp>2009-01-23T09:13:35.000+00:00</Timestamp>
  <Originator>IC Desk</Originator>
</AuditTrailEntry>
```

Deriving Cases from Enterprise Data

Having found a way to convert the enterprise data to the MXML format, cases need to be constructed. In view of compositional association and cross-document association, a coherent strategy for grouping instances needs to be developed so that the shifting focus between different semantically interrelated business objects is handled appropriately. A strategy is therefore pursued for grouping the entire interrelated chain of IC elements belonging to a complaint into one case. Important IC elements certainly include CRM activities, business partners, products, and documents. This strategy uses two types of information from the SAP CRM data model.

1. A set of CRM activities, or more precisely CRM activity instances.
2. A set of pairs establishing relations between two CRM activities or one CRM activity and one IC element. The relations between two CRM activities is expressed by a predecessor-successor relationship; the relationship between one CRM activity and one IC element by a causal relationship.

Since the document flow of the CRM system stores the predecessors, the successors, and the links between the CRM activities and relevant IC elements, the document flow can be used to trace CRM activities and other process-relevant elements upon which the business process is carried out. An example showing the document flow of CRM activities is seen in Table 3.5.

Six CRM activities are involved in the document flow, namely the CRM activities Cust. Relations 137, Cust. Relations 138, Cust. Relations 139, Cust. Relations 140,

Table 3.5.: Document flow in the SAP CRM system

Activity	Role	Document Type	Description	Date	Time
Cust. Relations 137	Incoming comm. participant	Activity CRM	Cust. Relations 138		16:30:38
	Comm. participant	Activity CRM	Cust. Relations 138		16:07:36
		Business partner	Michael Ross	13.04.2009	16:07:06
Cust. Relations 138	Successor	CrmAnchorObject	E2ADEE...2A67A8		16:05:53
		Activity CRM	Cust. Payments 141		16:30:38
		Activity CRM	Cust. Relations 138		16:07:28
Cust. Relations 139	Incoming comm. participant	Activity CRM	Cust. Relations 137		16:07:36
	Comm. participant	Activity CRM	Cust. Relations 137		16:30:38
	Successor	Activity CRM	Cust. Relations 139	13.04.2009	16:20:54
	Predecessor	Activity CRM	Cust. Relations 137		16:07:28
Cust. Relations 140	Predecessor	Activity CRM	Cust. Relations 138	13.04.2009	16:20:54
Cust. Relations 141	Successor	Activity CRM	Cust. Relations 142		16:35:48
	Comm. participant	CrmAnchorObject	H8TLJU...5A43Z8	13.04.2009	16:35:00
Cust. Payments 142	Predecessor	Activity CRM	Cust. Relations 137	13.04.2009	16:30:38
Cust. Relations 142	Predecessor	Activity CRM	Cust. Relations 140		16:35:48
	Comm. channel	Comm. participant	320004RAW320041	13.04.2009	16:36:01

3. Exploiting New Data Sources for Process Mining

Cust. Payments 141, and Cust. Relations 142 (also known as 137, 138, 139, 140, 141, and 142.).

To manage the document flow, a number of roles exist describing the interdependency of any two given elements. Note the following roles: successor, predecessor, communication participant, incoming communication participant, and communication channel. Role *successor* is used to express the semantic relation “creates”, while role *predecessor* represents a relation that can be best expressed with the semantics “is created by”. Role *communication participant* stands for semantics “changes” and role *incoming communication participant* indicates that the participant of the communication “is changed by”. Because of the symmetry of these relations – if activity A creates or changes activity B , then activity B is created or changed by activity A – for each predecessor a successor exists, for each incoming communication participant a communication participant, and vice-versa. The *communication channel* refers to the means of communication, such as e-mails, letters, and phone calls – the means is signaled with the semantics “is initiated by”. The document type describes the CRM element. Document type *CrmAnchorObject* denotes that the corresponding activity is the origin of a particular case. The document flow stores both the date on which and the time at which the communication took place.

Take the first document flow of activity 137 – it is changed by activity 138, which is itself changed by activity 137, both being participants of the *communication participant*. The role *successor* indicates that activity 137 has created activity 138 and activity 141.

The document flow depicted in Table 3.5 provides the IC elements $\mathcal{E} = \{137, 138, 139, 140, 141, 142, \textit{Ross}, \textit{mail}\}$, a relation of communication participants and incoming communication participants $\mathcal{R}_{cp/icp} = \{(137, 138), (138, 137)\}$, a relation of both predecessors and successors $\mathcal{R}_{p/s} = \{(137, 138), (138, 139), (137, 141), (140, 142)\}$, a relation of business partners $\mathcal{R}_{bp} = \{(137, \textit{Ross})\}$, and that of communication channels $\mathcal{R}_{cc} = \{(142, \textit{mail})\}$. The union over all relations provides the set of relationships $\mathcal{R} = \{(137, 138), (138, 139), (137, 141), (137, \textit{Ross}), (140, 142), (142, \textit{mail}), (138, 137)\}$.

The strategy to derive cases from the document flow is based on finding interrelated elements from the data as described above. That is, two elements are considered interrelated if and only if these ones are in the same case. An equivalence relation “belongs to the same case as” is constructed that specifies how to group the set of CRM activities and related IC elements so that every CRM activity of the set is exactly in one case, and the union of all the cases equals the original set. The grouping of single elements to cases corresponds to a partition.

The only disadvantage of this strategy is that elements that are created without reference to a CRM element will not be reconstructed as part of the case, thereby distorting the chain of IC elements – that said, processing of erroneous data is not a concern of this paper, and it is to assume that missing links are resolved in data preparation. Attention is now turned to the terms equivalence relation, equivalence class, and partition.

Definition 6 (Equivalence Relation)

Let \mathcal{R} be a binary relation defined over the set \mathcal{E} . \mathcal{R} is said to be an *equivalence relation* if and only if it is reflexive, symmetric, and transitive. Equivalently, for all a , b , and c in \mathcal{E} :

- reflexive iff $(a, a) \in \mathcal{R}, \forall a \in \mathcal{E}$,
- symmetric iff $(a, b) \in \mathcal{R}$ then $(b, a) \in \mathcal{R}, \forall a, b \in \mathcal{E}$,
- transitive iff $(a, b) \in \mathcal{R}$ and $(b, c) \in \mathcal{R}$ then $(a, c) \in \mathcal{R}, \forall a, b, c \in \mathcal{E}$. *

An example follows.

Example 8 (Equivalence Relation)

Two elements are in relation if they are in the same case. Consider a set of elements $\{E_1, E_2, E_3\}$. Every element is in the same case as itself (reflexive). If element E_1 is in the same case as element E_2 , element E_2 is then in the same case as element E_1 (symmetric). Finally, if element E_1 is in the same case as element E_2 , and element E_2 is in the same case as element E_3 , then element E_1 is in the same case as element E_3 (transitive). The defined relation is obviously an equivalence relation. ◇

An equivalence relation partitions a set into several disjoint subsets, called equivalence classes. All the elements in a given equivalence class are equivalent among themselves, and no element is equivalent to any element from a different class.

Definition 7 (Equivalence Class)

Let \mathcal{R} be an *equivalence relation* on a set \mathcal{E} , then *equivalence class* denoted by $[.]_{\mathcal{R}}$ is generated by the elements $y \in \mathcal{E}$ so that $[y]_{\mathcal{R}} = \{a \mid a \in \mathcal{E} \text{ and } (y, a) \in \mathcal{R}\}$. If the equivalence relation \mathcal{R} is clear from the context, the subscript is omitted. *

In this case, the notation $[a]$ describes the equivalent class of \mathcal{E} , that is, the case in which the element a is. A family of equivalence classes generated by the elements of \mathcal{E} defines a partition of set \mathcal{E} . Such partition is unique. Patently, equivalence classes of any two elements are either disjoint or equal, that is, $[a] \cap [b] = \emptyset$ or $[a] = [b]$.

Definition 8 (Partition)

Let \mathcal{E} be a set. A *partition* of \mathcal{E} is a set $\pi = \{C_i \subseteq \mathcal{E} \mid 1 \leq i \leq k\}$ of pairwise disjoint and non-empty subsets C_i of \mathcal{E} such that every element of \mathcal{E} is an element of exactly one of these subsets (or blocks) and that their union is equal to \mathcal{E} :

- $\forall i : C_i \neq \emptyset$,
- $\forall i \neq j : C_i \cap C_j = \emptyset$,
- $\bigcup_{1 \leq i \leq k} C_i = \mathcal{E}$. *

Returning to the document flow as depicted in Table 3.5, note that relation \mathcal{R} is not an equivalence relation – it fails to meet the requirements as specified in Definition 6. An operation on relation \mathcal{R} that consists of adding to it all the pairs successively induced by reflexivity, symmetry, and transitivity will resolve this difficulty.

The following description of the corresponding algorithm illustrates case construction from enterprise data.

Algorithm for Case Construction from Enterprise Data

Algorithm *BO2MXML* is a *coarsening* partition algorithm – it starts with the finest possible partition and gradually coarsens elements into blocks until the final equivalence classes are found. The partition is built on the equivalence relation “belongs to the same case as”. Algorithm *BO2MXML* considers only closed complaints – one is said to be closed only if all interrelated CRM activities contain the status *closed*.

This algorithm can be verbalized. Augment non-equivalence relation \mathcal{R} such that it becomes an equivalence relation – now reconstruction of the document flow of the IC elements can follow. The equivalence relation is induced by the reflexive, transitive, and symmetric closure of relation \mathcal{R} .

The above is seen in Example 9, which adopts the data from the document flow as shown in Table 3.5.

Example 9 (Closure of Relation)

Let $\mathcal{M} = \{137, 138, 139, 140, 141, 142, Ross, mail\}$ be a set of MXML events and $\mathcal{R} = \{(137, 138), (138, 139), (137, 141), (137, Ross), (140, 142), (142, mail), (138, 137)\}$ a set of relationships among those events. Since relation \mathcal{R} fails to be reflexive, symmetric, and transitive, algorithm *BO2MXML* successively adds all missing pairs with which relation \mathcal{R}' becomes an equivalence relation, that is, relation \mathcal{M} is closed.

$$\begin{aligned} \mathcal{R}' = \mathcal{R} \quad &\cup \quad \{(137, 137), (138, 138), (139, 139), (140, 140), (141, 141), \\ &\quad (142, 142), (Ross, Ross), (mail, mail)\} \\ &\cup \quad \{(139, 138), (141, 137), (Ross, 137), (142, 140), \\ &\quad (mail, 142)\} \\ &\cup \quad \{(137, 139), (139, 137), (138, 141), (141, 138), (139, 141), \\ &\quad (141, 139), (138, Ross), (Ross, 138), (139, Ross), \\ &\quad (Ross, 139), (141, Ross), (Ross, 141), (140, mail), \\ &\quad (mail, 140)\} \end{aligned}$$

Now, relation \mathcal{R}' satisfies reflexivity, symmetry, and transitivity. It follows that relation \mathcal{R}' is an equivalence relation. \diamond

The algorithm is expressed in pseudo-code – it is based on four phases: selection, conversion, transformation, and partitioning. Salient aspects follow.

In line 1, the algorithm enters the selection phase in which function `selectCRMActivities(GUID, \mathcal{E} , \mathcal{R})` is applied. This function first calls `BAPI BAPI_ACTIVITYCRM_GETDETAILMULT`, which provides information about the CRM activities by passing to it one or more CRM activity GUIDs. This BAPI returns details about header, business partners, reasons, dates, status, documents, products, and, most importantly, the list of relationships among the IC elements. Algorithm *BO2MXML* then uses this to prepare two lists of data: the set of IC elements $\mathcal{E} = \{E_i \mid 1 \leq i \leq n\}$ and the set of relationships among these elements $\mathcal{R} = \{(E_{i1}, E_{i2}) \mid E_{i1}, E_{i2} \in \mathcal{E}, 1 \leq i \leq m\} \subset E \times E$.

From line 2 to line 9 algorithm *BO2MXML* is in its conversion phase. Line 3 enters the first loop. In each iteration over the loop, function `addAll(createActivities(e))` generates single fine-grained activities from each IC element e in \mathcal{E} . This set of activities

BO2MXML Case construction of enterprise data

INPUT: GUIDs – identifier of CRM activities

OUTPUT: π – cases

```

1: list.selectCRMActivities(GUID,  $\mathcal{E}$ ,  $\mathcal{R}$ );
2:  $\mathcal{A} = \emptyset$ ;
3: for all  $e$  in  $\mathcal{E}$  do
4:    $\mathcal{A}$ .addAll(createActivities( $e$ ));
5: end for
6:  $\mathcal{M} = \emptyset$ ;
7: for all  $a$  in  $\mathcal{A}$  do
8:    $\mathcal{M}$ .add(createMXMLEvent( $a$ ));
9: end for
10:  $\mathcal{R}' = \text{transformRelation}(\mathcal{R}, \mathcal{E}, \mathcal{M})$ ;
11:  $\pi = \{ \{m_1\}, \{m_2\}, \dots, \{m_{|\mathcal{M}|}\} \}$ ;
12: for all  $(m, m') \in \mathcal{R}'$  do
13:   if  $[m] \neq [m']$  then
14:      $\pi$ .remove( $[m]$ );
15:      $\pi$ .remove( $[m']$ );
16:      $\pi$ .add( $[m] \cup [m']$ );
17:   end if
18: end for

```

is denoted as relation \mathcal{A} . In line 7, the algorithm enters the second loop during which each iteration function `add(createMXMLEvent(a))` converts the activities to the MXML format. The function returns relation \mathcal{M} consisting of pairs of MXML events.

The transformation phase of line 10 involves inferring of relationships between activities obtained from the relationships among the MXML events. The corresponding function `transformRelation($\mathcal{R}, \mathcal{E}, \mathcal{M}$)` transforms relation \mathcal{R} into \mathcal{R}' . Three arguments are passed to the function: a set of relationships \mathcal{R} , a set of IC elements \mathcal{E} , and a set of MXML events \mathcal{M} . Consider the IC elements $\{e_1, e_2, e_3\}$ and the relationship (e_1, e_2) . The MXML events $\{m_{11}, m_{12}, m_{13}, m_{21}, m_{22}, m_{31}\}$ reflect that IC element e_1 results in activities m_{11}, m_{12} , and m_{13} , IC element e_2 consists of activities m_{21} and m_{22} , and the last IC element e_3 corresponds to activity m_{31} . The relationships are obtained by applying the relationships of the IC elements to the activities $\{(m_{11}, m_{21}), (m_{11}, m_{22}), (m_{12}, m_{21}), (m_{12}, m_{22}), (m_{13}, m_{21}), (m_{13}, m_{22})\}$. The transformation phase ends with the reflexive, symmetric, and transitive closure of relation \mathcal{R} , as a result of which the function returns equivalence relation \mathcal{R}' .

The subsequent partition phase occurs in line 11 to line 18. Using a block for each element of relation \mathcal{M} as initial blocks the algorithm starts the partitioning with the finest possible partition in line 11. In line 12, a loop starts iterating over every pair of MXML events to coarsen gradually the initial blocks. Whenever a pair of IC elements is in direct relation (i.e., predecessor-successor or causal), this pair belongs to the same

3. Exploiting New Data Sources for Process Mining

case and needs to be assigned to the same block. Therefore algorithm *BO2MXML* allows one to see whether the blocks are different even if they belong together (line 13) and if so unites these different blocks of two MXML events as depicted in line 14 to line 16. After the last iteration, in line 20, the final partition π is found and is the set of all cases where each block in the partition is one case. The resulting cases do not have a chronological order. The order of the MXML log is constructed during the loading of the log into ProM with respect to time stamps.

Using the data from Example 9, the following example demonstrates the partitioning technique. For the purpose of brevity, the elements 137, Ross, mail, as well as the corresponding relations are excluded.

Example 10 (Partition)

Consider a set of elements $\mathcal{E} = \{138, 139, 140, 141, 142\}$ and an equivalence relation $\mathcal{R}' = \{(138, 138), (139, 139), (140, 140), (141, 141), (142, 142), (138, 139), (139, 138), (138, 141), (141, 138), (139, 141), (141, 139), (140, 142), (142, 140)\}$.

$$\begin{aligned}
\pi_0 &= \{\{138\}, \{139\}, \{140\}, \{141\}, \{142\}\} \\
(138, 138) \pi_1 &= \{\{138\}, \{139\}, \{140\}, \{141\}, \{142\}\} \\
(139, 139) \pi_2 &= \{\{138\}, \{139\}, \{140\}, \{141\}, \{142\}\} \\
(140, 140) \pi_3 &= \{\{138\}, \{139\}, \{140\}, \{141\}, \{142\}\} \\
(141, 141) \pi_4 &= \{\{138\}, \{139\}, \{140\}, \{141\}, \{142\}\} \\
(142, 142) \pi_5 &= \{\{138\}, \{139\}, \{140\}, \{141\}, \{142\}\} \\
(138, 139) \pi_6 &= \{\{138, 139\}, \{140\}, \{141\}, \{142\}\} \\
(139, 138) \pi_7 &= \{\{138, 139\}, \{140\}, \{141\}, \{142\}\} \\
(138, 141) \pi_8 &= \{\{138, 139, 141\}, \{140\}, \{142\}\} \\
(141, 138) \pi_9 &= \{\{138, 139, 141\}, \{140\}, \{142\}\} \\
(139, 141) \pi_{10} &= \{\{138, 139, 141\}, \{140\}, \{142\}\} \\
(141, 139) \pi_{11} &= \{\{138, 139, 141\}, \{140\}, \{142\}\} \\
(140, 142) \pi_{12} &= \{\{138, 139, 141\}, \{140, 142\}\} \\
(142, 140) \pi_{13} &= \{\{138, 139, 141\}, \{140, 142\}\} \\
\pi &= \pi_{13}
\end{aligned}$$

The initial partition π_0 includes a block for each element in the set of elements \mathcal{E} . The first iteration does not change the initial blocks because element 138 already has its own block. Partition π_1 becomes the partition for the next iteration. Iterations two to five, like iteration one, change nothing at the initial blocks. In the sixth iteration element 138 and element 139 do not share the same block. Since they are equivalent with respect to the equivalence relation, the blocks are replaced with block $138 \cup 139$. The algorithm then repeats the entire process. When every iteration has been generated, the resulting partition π_{13} gives the final equivalence relation π . The first case comprises elements $\{138, 139, 141\}$ and the second comprises elements $\{140, 142\}$. \diamond

The screenshot shows the BO2MXML application window. The title bar includes 'Program Edit Goto System Help' and the SAP logo. The main interface is divided into three sections: 'Data Selection', 'MXML Conversion', and 'Output'.

Data Selection:

- ☒ Complete cases
- Activity GUID: [] to []
- Activity type: ALL to []
- Role in document flow: [] to []
- Responsible Group: A11 to []
- Selection period of time: [] to []

MXML Conversion:

- ☒ MXML format
- Data Attribute: GUID to []
- System: SAP CRM system
- Workflow log adress 1: http://
- Workflow log adress 2: http://
- Workflow description: Complaint handling process
- Source program: BO2MXML
- Process ID: Air1
- Process description: Interaction Center (IC)

Output:

- Output file: c:\Documents\Air1.xml

Figure 3.5.: Implementation of algorithm *EPCIS2MXML*

A number of parameters control data selection – see Figure 3.5. The resulting MXML file can be opened in ProM for the purpose of applying process mining.

3.5. Conclusion

The empirical study of process mining provides a number of lessons as to the usability of RFID events and enterprise data – each to be discussed.

3.5.1. Process Discovery in Enterprise Systems

ERP systems, having diffused best practices, provide important information for re-engineering efforts. However, given current monitoring solutions are either platform-specific or necessitate manual collation of information spread across a variety of information systems and applications. Therefore, one seeks to investigate enterprise data exploitation to attain insight into how business processes are handled within an enterprise system.

3. Exploiting New Data Sources for Process Mining

The study of the enterprise system package CRM's data structure's complexity has shown, on surprisingly, that it complicates defining activities and cases. Therefore, collaboration with business objects is highly desirable. The standardized interfaces BAPI offer clear advantages over direct access through tables, keeping in mind that process mining requires, at bare minimum, knowledge of the relations among the business objects – depending on the data, a thorough knowledge of the data structure of the ERP package is strongly preferable.

The representation of the facts that are not covered by the enterprise transactions, though they ought to have been, has been explored. The refinement of transactions is crucial for comparing the activities with those of a reference model. Despite the perpetual problem of ordering the activities when using transactional data, the benefit of refining the CRM activities to the granularity of that of a reference model is, in many cases, well worth the inevitable degree of inaccuracy. This inaccuracy begs for the development of analytic tools, discussed in Chapter 6.

Finally, the problems of composition and association, discussed above, had presented a considerable challenge for making process mining applicable as to database-oriented data. These problems' investigation has shown that it is possible to reconstruct complaint processes even when there are compositions and associations among the CRM activities.

Herein has been presented an algorithm along with a prototypical implementation. This algorithm induces an equivalence relation from the document flow of the CRM activities to reconstruct cases of IC elements belonging to one complaint. The document flow, not restricted to the CRM system, is used in all SAP enterprise systems, as a result of which the partition approach can be used not only as to CRM activities and related business information but to a number of enterprise applications, such as purchasing and sales processing. The automated reconstruction of the document flow substantially reduces manual collocation. As a result, the application of process mining in enterprise systems naturally becomes more attractive to business. Especially noteworthy, regarding continuous process improvement, the following chapters will explain that the resulting process models provide valuable information. Bear in mind that the data selection is restricted to the proprietary application domain IC of the CRM system.

This work may provoke further research from various perspectives. More will be necessary to understand fully composition as a vital aspect to develop comprehensive solutions for dealing with associations within and among business objects. Constructive data preparation operations, such as the construction of derived activities, should be investigated in more detail – the study of intermediate documents and business documents used for the standardized communication among enterprise packages and other information systems might be helpful.

3.5.2. Process Discovery in RFID Environments

With the rise of RFID and the ubiquity of computing devices, far more events are observed and available for process analysis. At this writing, most RFID scenarios are implemented to support operational requirements of organizations while strategic potentials are somewhat neglected (Ivantsynova, 2008). Therefore, to unleash the full

potential of process mining, it is necessary to avail oneself of opportunities associated with the RFID technology's diffusion by analyzing information and product flow in tandem. Due to the current gap between raw data and process analysis capability, this paper has contributed the domain of RFID and process mining research.

Current research challenges to making RFID event logs accessible to process mining analysis (case identification and focus shifts) have been addressed. It has been shown that the EPCglobal standard for processing RFID can be used to construct cases from such events. In a word, an algorithm has been proposed to group EPCIS events to cases. It deals with various types of events including assembly and disassembly ones. It has been implemented as part of a tool that converts EPCIS event logs to MXML files, which can be processed with ProM.

From the perspectives of both business and research, questions remain as to properly supporting the shifting focus. This support is likely to include the construction of a tree of electronic product codes belonging to the same instance in such a way that the most coarse-grained physical object is on top and with all childEPCs included. In a supply chain, a number of business entities might be interested in following the event stream on different levels of granularity. Therefore, future endeavors should include research both in mining algorithms and in interactive tool support.

4. Process Discovery with Process Mining

This chapter investigates opportunities and challenges emerging from using process mining for real-life processes.

Process-unaware data is to a great extent unexploited for the use of process mining. This chapter deals with the dramatic possibility of exploiting this data to achieve greater insight into processes handled within RFID and enterprise environments. And this chapter elucidates the inter-organizational use of process mining – an environment as to which little has been gleaned due to missing or inconsistent process identifiers. Finally, concern is shown as to mining workflows in a new domain, especially as to ITSM process improvement.

The case study method has been chosen as a qualitative research method because of the research’s exploratory nature – it facilitates analyses of contemporary phenomena in a real word context. And this method is a means of collecting and analyzing data to gain a broad and deep understanding of a given situation. It is particularly useful for both investigations as to which research and even theory are inchoate and those as to which a seemingly sound theory may be inadequate or inapposite (Benbasat et al., 1987). The case research method is therefore well-suited to capture practitioners’ knowledge and to develop theories from it. Of course, extrapolation from case-based analyses suffers from the intrinsic unreliability of generalizing from small samples – that said, the cases herein have been considered deeply (Flyvbjerg, 2006). A diverse case sampling strategy allows the most representative cases to be chosen, a strategy likely to produce apt sample cases (Gerring, 2007, p. 100).

Consequently, cases have been chosen that vary as much as possible as to their dimensions – these cases answer the following questions:

1. Are process mining techniques suitable for deriving as-is process models for the new data sources RFID events and ERP data?
2. What challenges and obstacles exist in the context of process mining, reference models, and process improvement?

Section 4.1 presents the case studies. Section 4.2 applies process mining to them. Concluding, Section 4.3 discusses the results.

4.1. Case Studies

The case studies were conducted between July 2008 and April 2010. The participating organizations are noted in Table 4.1. Two organizations requested anonymity. All

4. Process Discovery with Process Mining

organizations are headquartered in Europe. The number of employees ranged from approximately 15 to over ten-thousand employees.

Naturally, the type of data gathered depended on the research questions. The following provides key information, which is summarized in Table 4.2.

1. **Sector.** Similar processes in different industries can differ widely. Take a customer interaction. A customer wishes to interact with a telecommunications provider – this typically occurs directly as the product or service is used. In contrast, in the airline industry complaints are virtually always made after a given flight is over.
2. **Data Source.** The sample selection focuses on various types of information systems, such as data-centric systems (e. g., CRM system) and process-centric information systems (e. g., WFMS).
3. **Data Type.** The sample selection is rich – it contains a variety of sources of information ranging from single events to transactional data.
4. **Product and Information Flow.** Improving processes requires linking information and product flow. Note that both the flow of information and the flow of products are often nonlinear and multi-directional – information direction might, for instance, be either opposite or parallel to the product flow.
5. **Process Types.** These cases encompass a great deal of variation in process types because IT processes and business processes are both considered.
6. **Products and Services.** The sample selection contains processes with different products and services in order to observe their influence on process improvement.
7. **Inter-Organizational Use.** The significance of the sample selection lies in the fact that some cases involve a series of organizations rather than a single organization.

4.1.1. Car Production and Delivery

Consider a hypothetical situation⁴⁵.

The simulator used for the generation of the EPCIS events is called the *Supply Chain Editor* and is a means of modeling and simulating the product flow in supply chains. This editor has been devised as part of the Building Radio Frequency Identification for the Global Environment (BRIDGE) project, about which more in due course. The editor produces RFID events that are realistic in an actual supply chain. Naturally, all RFID events follow the EPCIS standard in format and content. The RFID events are stored in the open source EPCIS Fosstrak⁴⁶, which implements the EPCglobal Network specification.

During runtime the editor simulates numerous product instances, which enter the stream, and are then routed through the supply chain. During this routing, the products

⁴⁵ This work is published in Gerke et al. (2009c).

⁴⁶ <http://www.fosstrak.org>

Table 4.1.: Organizations of use cases

ID	Sector	Role	Company	Head- quarter	Local	Employees Worldwide
Air-1, Air-2	Airline	Passenger airline	—	DE	900	900
Pha-1	Pharma- ceutics	Manufacturer	Athlone Laboratories	IL	85	85
			Actavis	UK	450	10,500
			Sandoz	UK	72	23,000
		Packer	Tjoapack	NL	150	n.a.
		Distributor	Kent Pharmaceuticals	UK	125	250
			CPG Logistics	UK	139	n.a.
		Logistic service provider	Healthcare Logistics	UK	15	15
		Wholesaler	UniChem	UK	26	26
Aut-1	Automotive	Hospital	Bart and The London NHS Trust	UK	7,800	7,800
		OEM				
		1 st tier supplier	simulated			
		2 nd tier supplier				
Tel-1 – Tel-4	Telecom- munications	Logistic service provider				
		Internet service provider	—	DE	81,336	259,920







Table 4.2.: Properties of use cases

ID	Data Source	Data Type	Flow	Process	Product/ Services	Cross- company	Period ⁴⁷	Number	
								Events ⁴⁸	Cases
Air-1	CRM	transactional	information	business	complaints		01/08-03/09	44, 006	4, 650
Air-2	system						01/08-08/08	18, 536	2, 340
Aut-1	RFID	RFID events	product & information	business	vehicle doors	✓	22 hours	41, 534	8, 573
Pha-1	RFID	RFID events	product	business	drugs	✓	10/07-05/08	363, 688	31, 766
Tel-1	WFMS	process-aware	information	IT	e-mail		06/09-08/09	8, 879	797
Tel-2					S_1		06/09-11/09	20, 469	1, 816
Tel-3	WFMS	process-aware	information	IT	S_2		06/09-11/09	49, 452	4, 182
Tel-4					S_3	✓	07/09-12/09	64, 593	6, 070

⁴⁷ Period of data selection⁴⁸ MXML events

may pass various spatial locations (hereafter “nodes”) in which events are generated (e. g., the movement of goods). The nodes mimic RFID readers that might be found in sites, such as warehouses, plants, and distribution centers. The various types of nodes are illustrated in Table 4.3.

Table 4.3.: Behavior of nodes

Producer	Observer	Assembler	Packer	Unpacker	Fabricator
					

The *producer* manufactures a given quantity of items of a product. The *observer* observes a given number of items. The *assembler* assembles a given quantity of items of different products to produce a single item of a product. The editor allows one to differentiate between production and shipment aggregation. The *packer* produces an item of a type product in quantity. The *unpacker* unpacks items of a product type. The *fabricator* transforms a crude product into a refined product. All this behavior can be associated with or disassociated from business transactions as to all processed products. The behavior can be elucidated in greater detail using fields such as quantity, product, and delay. Table 4.4 shows the events generated based on nodes’ behavior.

The design of the automobile supply chain in case Aut-1 stems from both M2 and D2 scenarios according to the SCOR model (Supply Chain Operations Council, 2006). The adoption of this reference model produces production and delivery processes organizations are expected to execute in an actual business situation. Figure 4.1 illustrates the resulting supply chain model – the dotted line separates the M2 process from the D2 process. SCOR abbreviations – Source Make-to-Order (S2), M2, D2, as well as EM (enable activity) – are employed. The production of modular door systems involves a second-tier supplier (S2), a first-tier supplier (M2.1 - M2.6, D2:1 - D2:10), an Original Equipment Manufacturer (OEM) (D2:13), and a shipper (D2:12). The systems integrate window regulator, door control unit, loudspeaker, wiring harness, and the complete latch module onto a carrier base plate.

Table 4.4.: Events generated based on behavior

Behavior	EPCIS Event Type	Event Field
Producer	ObjectAdd	\emptyset
Observer	AggregationObserve or ObjectObserve	childEPCs
Assembler	ObjectAdd and ObjectDelete, AggregationAdd	childEPCs
Packer	AggregationAdd	childEPCs
Unpacker	AggregationDelete	childEPCs
Fabricator	ObjectAdd and ObjectDelete	sourceEPC

4. Process Discovery with Process Mining

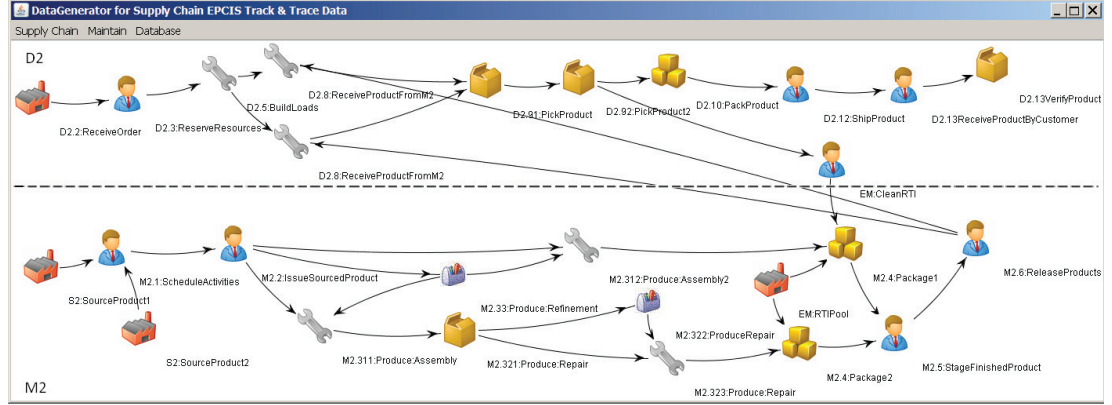


Figure 4.1.: Automobile supply chain

A make-to-order environment is one in which products are manufactured only in response to a customer order. The manufacturing process adds value to products through mixing, separating, and forming (Supply Chain Operations Council, 2006). The following description of the M2 process assumes that it is supported by RFID technology. All captured tag data is categorized by event types ObjectEvent, TransactionEvent, QuantityEvent, and AggregationEvent.

The scenario starts with the coordination of all intermediate production activities prior to the scheduling of the operations to be performed in making modular vehicle doors (M2.1:ScheduleActivities). The scheduling of production activities results in a production plan authorizing the production of a certain quantity of a product. According to this plan, the sourced products are selected and physically moved from a stocking location to a specific point in the production area (M2.2:IssueSourcedProduct). The products' movement is monitored and recorded by means of ObjectEvents at specific sites, such as doorways and shelves. In the production area, activities are performed to convert the products from a semi-finished state to a state of completion (M2.3:Produce)⁴⁹.

The completed products, the doors in question, are immediately tagged with an EPC. The related ObjectEvent contains information about the EPC, the location, and the time. The doors are then sequenced onto returnable items (M2.4:Package). The packaging step is linked to an AggregationEvent and a TransactionEvent, which in turn link the returnable item (RTI), the doors on the RTI, and the related production plan. The returnable items are allocated from an RTI pool (EM:RTIPool). They carry the doors until activity D.2.92 after which they are replaced with OEM specific transportation means. The cleaned returnable items (EM:CleanRTI) are returned to the packaging

⁴⁹ The production is represented by the two assembly operations M2.311:Produce:Assembly and M2.312:Produce:Assembly2 (i.e., product loudspeaker and product plate), the three repair activities M2.321:Produce:Repair (i.e., disassembly of faulty loudspeaker), M2.322:Produce:Repair (i.e., repair of faulty loudspeaker), and M2.323:Produce:Repair (i.e., assembly of intact loudspeaker), and one refinement activity M2.33:Produce:Refinement (i.e., product galvanized plate) in the simulator.

site. Each move of the returnable items is captured by means of ObjectEvents at the packaging site doors. The sequenced doors are moved into a temporary holding location to await movement to a finished goods location (M2.5:StageFinishedProduct). The doors' storage is monitored by means of QuantityEvents. Finally, testing is conducted and release documentation prepared (M2.6:ReleaseProducts). The doors are linked to the production plan via a TransactionEvent.

Appendix E contains an in-depth D2 process description.

4.1.2. Pharmaceutical Drug Delivery

The project BRIDGE was created to facilitate the deployment of RFID applications in Europe (Jenkins et al., 2007a). The *Pharma Traceability Pilot* was established to track and trace a drug's production within the pharmaceutical supply chain. The *pilot* operated in a live production environment, that is, real products were supplied, with technologies and standards then available, for hospital use (Jenkins et al., 2008). The key elements were the use of an implementation of the EPCglobal Network for supply chain wide data collection and the use of RFID tags on all levels of product packaging, such as items, cases, and pallets. Figure 4.2 illustrates the supply chain during which drugs were tracked at every point from the manufacturers to the hospital pharmacy. This use case, as to which a number of names and locations are abbreviated, is hereafter “Pha-1”. Athlone Laboratories (Athlone), Actavis, and Sandoz are the manufacturers. Athlone's antibiotics are manufactured after receiving an order of Kent Pharmaceuticals (Kent). In response to the fact that antibiotics cannot be packed under the same conditions as other pharmaceutical products, Athlone undertakes the serialization and packaging at its site in Ireland. The drugs, in cases, are delivered to Kent on pallets and

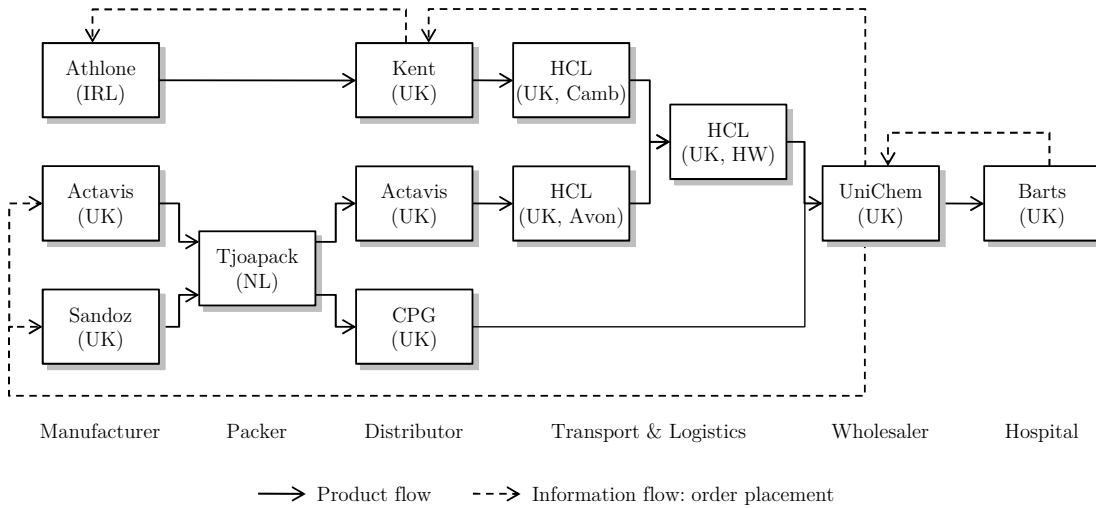


Figure 4.2.: Pharmaceutical supply chain

4. Process Discovery with Process Mining

held in storage awaiting a UniChem order. Similarly, Sandoz and Actavis manufacture and pack various drugs according to the production schedule triggered by a UniChem order. They ship these drugs products to Tjoapack where serialization is added to the packaging, after which they are shipped in cases and on pallets to storage locations, namely CPG Logistics (CPG) and Actavis's warehouse. Actavis plays overlapping roles in the drug delivery process since the manufacturer Actavis distributes its own products. The distributor Kent performs this task on Athlone's behalf. CPG acts as the distributor for Sandoz products – CPG unlike Kent, does not take title of the goods but manages logistics based on Sandoz's instructions. Following UniChem orders, CPG delivers Sandoz products directly to UniChem. Movements of products for both the above stages are in bulk, that is, the products are either in cases or on single-product pallets. Based on a UniChem order, products stored at Kent and Actavis are cross-docked and transported to UniChem by Healthcare Logistics (HCL), at the site of which these products are stored awaiting orders of Bart and The London NHS Trust (Barts). Actavis products are moved to HCL's Avonmouth (Avon) depot for cross-docking and on-shipping on their way to their High Wycombe (HW) depot, while Kent products are transported to HW via their Cambridge (Camb) depot. The shipment is received at the wholesaler UniChem, which has purchased the pharmaceutical products from the manufacturers. The hospital pharmacy Barts is the penultimate step after which the medication reaches the patient.

The underlying process has been broken into three processes – receipt, picking, and dispatch – to illuminate the wholesaler's event generation.

The pallets, mostly mixed product pallets, are received into UniChem's goods inwards by scanning the pallets off the incoming vehicle to reveal their content. The vehicle is allocated a unique code number GRAI that is scanned and associated with the pallets. The operator uses a hand held device to scan both the SSCC on the pallet and the GRAI on the truck. The pallets are then unloaded and the cases are stored in the goods receipt area. Finally, the operator stores the products – in this pharmaceutical example onto shelves (Lilley et al., 2008). The data is stored locally and is passed via the Internet to the EPCIS as a receipt event (Lilley et al., 2008). During the picking process the pallets are unpacked and disaggregated based on Barts instructions. This unpacking process is signaled to EPCIS as *AggregationEvents* with action field set to *Delete*. The product is picked to order and placed in returnable boxes for delivery to Barts. Each box is identified using a unique GRAI. Prior to being placed in the box, each product is scanned. Once the order is complete, the box is closed and sealed with a label representing the shipment's SSCC as to which either the box is the entire shipment or is part of an overall shipment. This process has items bearing SGTIN information (i.e., patient pack level) aggregated into containers bearing GRAI information and they in turn are aggregated into shipments denoted by an SSCC. An SSCC identifies shipment included one or more boxes (Jenkins et al., 2008). At the point of dispatch each box loaded onto the vehicle is recorded by scanning the GRAI on the vehicle and the SSCC on the box. Once the data is collected, the interface with the appropriate EPCIS, signals the event of goods dispatched (Lilley et al., 2008). The boxes are delivered to the Barts site.

Further information as to the products and the supply chain is available in the publicly accessible project deliverables (Jenkins et al., 2008; Lilley et al., 2008; Jenkins et al., 2007a,c,b,d) and on the web pages of the project (The BRIDGE Project, 2009).

4.1.3. Complaint Handling

The airline industry is strongly subject to the influences of customer perception – customer loyalty and retention are part and parcel of a successful airline. Good customer relations absolutely depend on the proper management and resolution of complaints. They have considerable potential for business process improvements because they give clues about risk. Identifying and reading these clues help an organization to fathom customer behavior. Take missed flights – complaints about them may indicate that check-in counters ought to remain in operation longer before an aircraft actually starts.

Complaints as a means of communication need to be integrated with core business processes else customer data is incomplete within different departments of the organization. It is therefore essential to implement structured methods for the efficient management of complaints. Proper complaint handling surely contributes to customer satisfaction and therefore to customer retention. Properly done, complaint handling is simplified, internal reaction time is reduced, processing status becomes more transparent, and underlying risks are identified very early.

The use cases *Air-1* and *Air-2* study a German passenger airline’s customer complaint handling process. The two use cases’ identifiers point at two different time intervals’ analyses as depicted in Table 4.2.

To foster communications, the airline has a call center in which agents deal with customers. Each customer contact – a service, an inquiry, or a complaint – is captured and organized by the CRM software solution Interaction Center (IC) from SAP (see Figure 4.3). As to each communication occurrence, the agent creates an interaction record that documents the processing in detail. Each interaction is registered as an activity. Each interaction is either be inbound or outbound, that is, either the customer initiated the communication or the agent did. If a customer has no record yet, the complainant is recorded as a new business partner and an activity of type *Customer Relations*, representing all interactions between customers and the airline, is created. It is possible to provide a solution to the complaint based on a predefined regulation procedure, an activity of type *Communication Operation* is executed, a response letter to the customer, now a business partner, is initiated and the matter is settled. If more detailed processing is necessary, an activity of type *Customer Relations, Lost & Found*, or *Customer Payment* is established based on the complaint’s grounds. For example, complaints about lost baggage are processed by the *Lost & Found* department. The transfer to this department takes place by generating a subsequent activity of type *Lost & Found*. The activity of type *Customer Payment* relates to any payment to or from a customer. If inter-departmental processing becomes necessary, the complaint is forwarded to the proper department through the creation of a sequential activity in the system. Figure 4.4’s Petri net illustrates inter-departmental processing of a complaint.

4. Process Discovery with Process Mining

The screenshot displays the SAP Interaction Center WebClient interface. At the top, the user is identified as John Taylor, Media Store. A navigation bar includes buttons for Accept, Reject, Hold, Retrieve, Hang Up, Dial, Transfer, Warm Transfer, Consult, Conference, Toggle, End, Dial Pad, and Next Call. The status is set to Ready. Below this, a search bar shows 'Complaint 0 *'. The main area is divided into two sections: 'Complaint Overview' and 'Header Details'. The 'Complaint Overview' section has tabs for Overview, Actions, Follow-up, Change History, and Document Flow. It contains fields for Reference Transaction, Description, Priority / Status (set to High and Open), Product ID, Reason, and Subject. The 'Header Details' section has tabs for Notes, Partners, and Business Context. It includes a 'Choose Text Type' dropdown set to 'Complaints Text' and a 'With Language' dropdown set to 'English'. At the bottom, there is an 'Item List' table with columns for Item, Item Type, Higher-Level Item, Product ID, Description, Quantity, Unit, and Status. The table is currently empty.

Figure 4.3.: Application Interaction Center

Each department captures the process of communication between customer and agent with details including the complaint's grounds, a description, and an external reference. Interactions are stamped with the name of the agent who created the activity, and this creation's date and time. A process status indicates how much of the activity has been completed. Furthermore, every activity is assigned a priority. During this communication, the department's agent attempts to placate or compensate the customer. This can lead to the creation of additional objects, such as business transactions (e. g., sales order and product), e-mails, and notes. At the communication's conclusion, all these objects are linked to the interaction record – naturally, everything that has happened during the interaction can be reviewed later on.

4.1.4. Service Operation

Dealing with unplanned interruptions to IT services – including failures, questions, and queries reported by users, technical staff, or those automatically detected and reported by event monitoring tools – is crucial for the success and efficiency of an Internet service provider (ISP)⁵⁰. Therefore, incident management strives to restore degraded or disrupted services to users as quickly as possible to minimize business impact.

⁵⁰ An ISP represents a special variant of an IT service provider that offers its customers' access to the Internet.

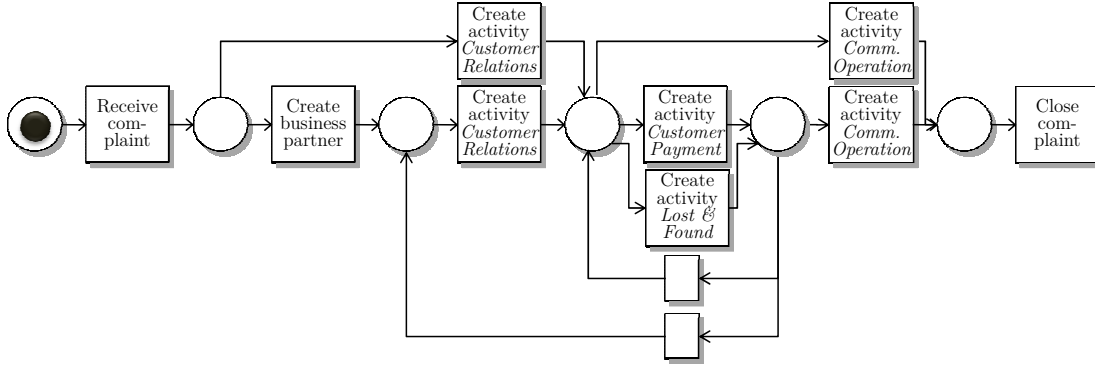


Figure 4.4.: Complaint handling process of a passenger airline

The service management of a German ISP for its IT service production is now analyzed. The ISP manages incidents, service requests, and communication with users via a service desk. After a service request has been reported to the service desk, a ticket, generated in the WFMS, is initially handled through the incident management process. This ticket passes through various processing steps until either the incident is disposed of or the problem is solved, at which time the ticket is closed. In general, the processing consists of the steps *Receive incident*, *Categorize incident*, *Analyze incident*, *Resolve incident*, and *Close incident*.

During ticket flow, the WFMS stores information of the current processing status as well as the corresponding time stamps. In addition, the groups dealing with incident handling fully document the details of any actions taken as to incident resolution, such as the originator of the action, affected service, relevant product, description of the incident, priority, and solution statements. Each processing step in the WFMS produces an entry in the “history of action”, such as the entries shown in Table 4.5.

Take the first processing step. Incident T1596654 was *forwarded* to the *1st level processing* with *Low* priority at 09:47:43 on 10 June 2009. This incident belongs to the process *Retrieve e-mail*, which is executed by the *Service desk*.

From a large set of services the services *e-mail*, S_1 , S_2 , and S_3 (henceforth use cases Tel-1, Tel-2, Tel-3, and Tel-4) are selected. The type of services S_1 , S_2 , and S_3 are hidden due to nondisclosure agreements. The services embrace two aspects (hereafter “differentiators”). First, the routing of the incidents among staff members involves numerous support groups. Second, the underlying ITSM processes’ complexity varies with the extent of the collaboration with external supply chain partners. Say that ISP offers the service S_3 , and therefore engages the service of a carrier – incidents clearly can be caused either by the ISP or the carrier. It follows that service S_3 is the most complex service.

Table 4.5.: Fragment of the original history of action

ID	Time Stamp	Activity	Process	Type	Priority	Action	Originator
T1596654	10.06.2009 09:47:43	1st level processing	Retrieve e-mail	Failure	Low	forwarded	SD ⁵¹
T1596654	10.06.2009 09:48:55	Assigned incident	Retrieve e-mail	Failure	Low	forwarded	SD
T1596654	10.06.2009 09:50:07	Assigned incident	Retrieve e-mail	Failure	Low	assigned	SD
T1596654	10.06.2009 11:33:56	Assigned incident	Retrieve e-mail	Failure	Low	in progress	SD
T1596654	10.06.2009 11:44:03	Additional info re-quested	Retrieve e-mail	Failure	Low	forwarded	SD
T1596654	10.06.2009 11:45:13	Additional info re-quested	Retrieve e-mail	Failure	Low	in progress	SD
T1596654	10.06.2009 15:51:09	Quality assurance	Retrieve e-mail	Failure	Low	forwarded	SD
T1596654	10.06.2009 15:51:45	Quality assurance	Retrieve e-mail	Failure	Low	forwarded	BOff ⁵²
T1596654	10.06.2009 15:52:49	Quality assurance	Retrieve e-mail	Failure	Middle	forwarded	BOff
T1596654	11.06.2009 09:21:41	Quality assurance	Retrieve e-mail	Failure	Middle	in progress	BOff
T1596654	11.06.2009 09:22:04	Incident closure	Retrieve e-mail	Failure	Middle	closed	SD
T1596654	11.06.2009 09:23:22	Incident closure	Retrieve e-mail	Failure	Middle	completed	SD

⁵¹ Service Desk⁵² Back Office

4.2. Process Discovery

This section looks at process discovery in business practice – it outlines the practical significance of process mining in a number of case studies – and it proves the concepts introduced in Chapter 3. The ProM framework as already introduced in Section 2.5.4 has been selected.

The general mining approach described here has been applied to all use cases, the data of which initially had to be made accessible to process mining. Depending on the use case, either algorithm *EPCIS2MXML* or algorithm *BO2MXML* has been employed. The resulting MXML file has been loaded in ProM and has been processed by the *Heuristics miner* (Weijters et al., 2006) to evaluate the results of the conversion process. This process mining algorithm provides threshold values that can count for infrequent behavior. The parameter *positive observations*, for instance, indicates how often two activities have occurred in relation to each other to establish a relationship (or the absence thereof) between them in the process model. The *Heuristics miner* has been applied with default parameter setting. It constructs a process model (see, for example, Figure 4.7) (henceforth model “ M_3 ”), with each node representing an action entry expressed with current status and associated frequency of occurrence and each arc a dependency between the action entries. The number inside the nodes is how many times an action has been executed. The upper number on the arcs shows the strength of a dependency relation between two activities. The closer the value is to 1, the more reliable the dependency relation between the connected activities is likely to be. The lower number is how often the relationship has been observed in the event log. Since each event log contains various start or end events, or both, artificial events (*ArtificialStartTask* and *ArtificialEndTask*) are generated to indicate specific start and end points for the process.

The ProM plug-in *Conformance Checker* (Rozinat and van der Aalst, 2008) provides the Token-Based Fitness (fitness) measure⁵³ to indicate the quality of the generated process model. This fitness metric evaluates how accurately the process model covers the observed actions in the event log by replaying the actions in it. Whenever a parsing error occurs, the error is registered and the parsing continues. The fitness value ranges from 0 to 1. The closer the value is to 1, the better the process model’s quality. The absence or paucity of reference values for the fitness obscures the classification of process models – that is, deeming them to be good or bad event log’s representatives is difficult or subject to dispute. Therefore absolute numbers of the fitness allow only for preliminary assessments of the models’ qualities. Section 4.3 will establish relationships among the single numbers to derive concluding assessments.

Proof of concept ensures that the requirements noted in Chapter 3 as to new data sources are satisfied. Finally, this section shows how organizations might improve their processes from process mining using a data source well-known in the process mining research field, namely workflows.

⁵³ This fitness metric is explained in detail in Section 6.3.1.

4.2.1. Car Production and Delivery

As to the requirements described in Sections 3.2.2 and 3.3.2, this section validates algorithm *EPCIS2MXML* with data from use case Aut-1 in a two-step approach.

The first step ensures that the MXML format provides a structure that can handle the life cycle of EPCIS events. The supply chain simulator noted above (Section 4.1.1) provides the two pieces of information needed to do so: the pre-established supply chain model (Figure 4.1) and the generated EPCIS events. The former is the to-be model of the supply chain (henceforth “ M_2 ”) because it describes the expected conversion result. The latter are events generated on the basis of model M_2 . These events are well-suited to validate the algorithm because they are noise-free, reproducible, and less complex than operational data.

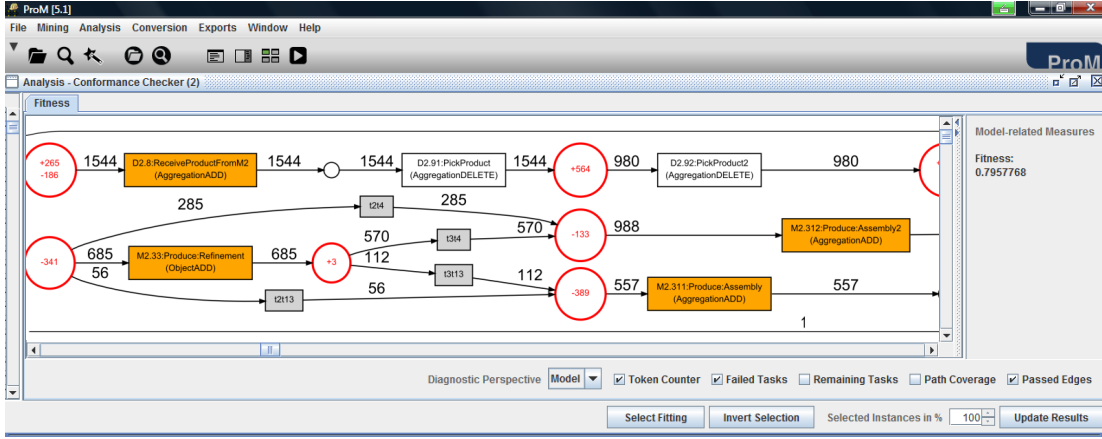
The second step ensures that model M_3 representing the output of the algorithm corresponds to model M_2 . The EPCIS events generated from the simulation model are processed by algorithm *EPCIS2MXML* with granularity setting *coarse-grain*. The resulting MXML file serves as a starting point for process mining.

Requirements R-1 and R-2: Correctness and Completeness

For the first-step of the validation, the ProM plug-in *Conformance Checker*, which aids in comparing the event log with model M_2 , is used. This comparison indicates whether the algorithm’s output is the expected result even if aggregation and disaggregation operations are included. Model M_2 is manually converted to a Petri net prior to this comparison. This net and the event log are fed into the plug-in *Conformance Checker*.

This analysis plug-in provides the fitness metric with which the correctness and completeness of the outcome of algorithm *EPCIS2MXML* can be proven. It also shows where discrepancies occur. Note that only completed instances have been considered, as a result of which the number of events has decreased. Figure 4.5 depicts the assessment of the comparison between the event log and model M_2 . The number of edges denotes the number of executions. Positive values in the circles show how often activities that should have been executed have not been; negative values there show how often activities have been executed though these executions were unplanned. For example, in contrast to model M_2 , 341 of 685 events passed activity *M2.33:Produce:Refinement (ObjectAdd)*. Since only 682 events were observed after the execution of activity *M2.33:Produce:Refinement (ObjectAdd)* – obviously three event observations were missing. The fitness yielded approximately a value of .80. It falling below the maximum is explicable. Logic points to the fact that the assessment’s negative and positive values – its deviations can be accounted for with reference to aggregation and disaggregation operations. Consider the refinement of the plate. This activity is observed both by an ObjectAdd event of the, say, galvanized plate, and by an ObjectDelete event in the sourceEPC of the unrefined plate. Consequently, in contrast to model M_2 , the event *M2.33:Produce:Refinement* occurs twice in the event log.

Having mined model M_3 from the event log, the second step of the validation is to compare this model with model M_2 , a comparison vis-a-vis transitions, quantities, and

Figure 4.5.: Validation of algorithm *EPCIS2MXML*

reliabilities. This step verifies that the discrepancies found in the previous validation step can indeed be explained with reference to aggregation and disaggregation. Figure 4.6 illustrates the event log opened in ProM from which model M_3 (Figure 4.7) is derived. Note that this model, for clarity sake, does not include TransactionEvents, about which more in due course.

Since model M_3 contained instances that stuck in the process⁵⁴ (in blue in Figure 4.7), another process model was derived consisting only of completed instances. This model (hereafter “ M'_3 ”) is shown in Figure 4.8. The fitness of model M'_3 with its event log was .99. This high value proves that model M'_3 corresponds perfectly with model M_2 .

Focusing attention on Figure 4.1’s depiction of model M_2 , one is certainly likely to conclude that model M'_3 is extremely similar because the left branch matches the D2 process and the right branch matches the M2 process. Specifically, all activities of model M_2 have been reconstructed in model M'_3 . Nevertheless, there are three deviations: additional activities, suspect quantities, as well as supplementary events, such as start events (in red in Figure 4.8) and end events (in blue in Figure 4.8). The following analyzes transitions, quantities, and transition reliabilities. Consider Figure 4.9, which juxtaposes an excerpt of model M'_3 (left-hand side) and one of model M_2 (right-hand side).

Begin with the issue of the sourced products plates and loudspeakers. The loudspeakers are assembled in *M2.311:Produce:Assembly* without an additional production activity and are represented by the transition going from *M2.2:IssueSourcedProduct* to *M2.311:Produce:Assembly*. The plates, however, need to be galvanized in the process *M2.33:Produce:Refinement* before they can be assembled together with the loudspeakers into the modular door system. The assembled doors then move to the activity *M2.321:Produce:Repair*.

⁵⁴ This will be explained in due course at which time the quantities will be verified.

4. Process Discovery with Process Mining

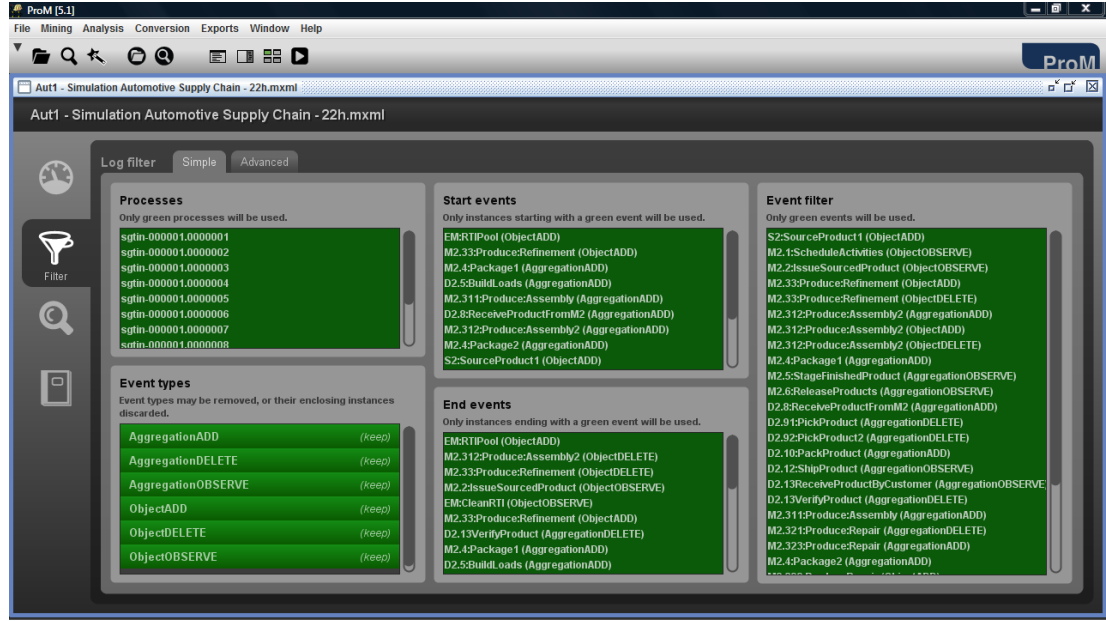


Figure 4.6.: Event log opened in ProM

To verify supplementary events, focus shifts must be examined. Whenever a new object is introduced by either an *AggregationAdd* event or an *ObjectAdd* event, a focus shift takes place, and new EPCIS events enter the product flow and accompany that product's events on the way to their destination. Take the assembled doors the life cycle of which begins in activity *M2.311:Produce:Assembly* – they are shown by a new start event in model M'_3 (in red). Each of the red arcs in Figure 4.8 and Figure 4.9 can be explained in this way. When objects fall into decline by either an *AggregationDelete* event or an *ObjectDelete* event something similar happens. See activity *M2.3211:Produce:Repair* in which the doors are disassembled for purpose of repair. At this point, faulty loudspeakers reach saturation and leave the product flow. The decline of this object is reconstructed as an end event in model M'_3 (in blue). The same is true for all blue arcs in Figure 4.8. Explanations for the additional activities are closely related to those for the supplementary events. A focus shift is often expressed in more than one event type (i.e., event type *Object* and *Aggregation*): one object going out of focus (i.e., action *Delete*) and one object coming into focus (i.e., action *Add*). In this regard, a plate's refinement is exemplary. In view of the above, one or more than one events of model M'_3 can relate to one activity of model M_2 .

The quantities can now be verified. As can be seen in activity *ArtificialStartTask* of model M'_3 , 8,573 events were processed, this number being the number of EPCIS events produced by the supply chain simulator during a run of twenty-two hours. Consider activity *M2.311:Produce:Assembly* in which a door handle and a galvanized plate were

4.2. Process Discovery

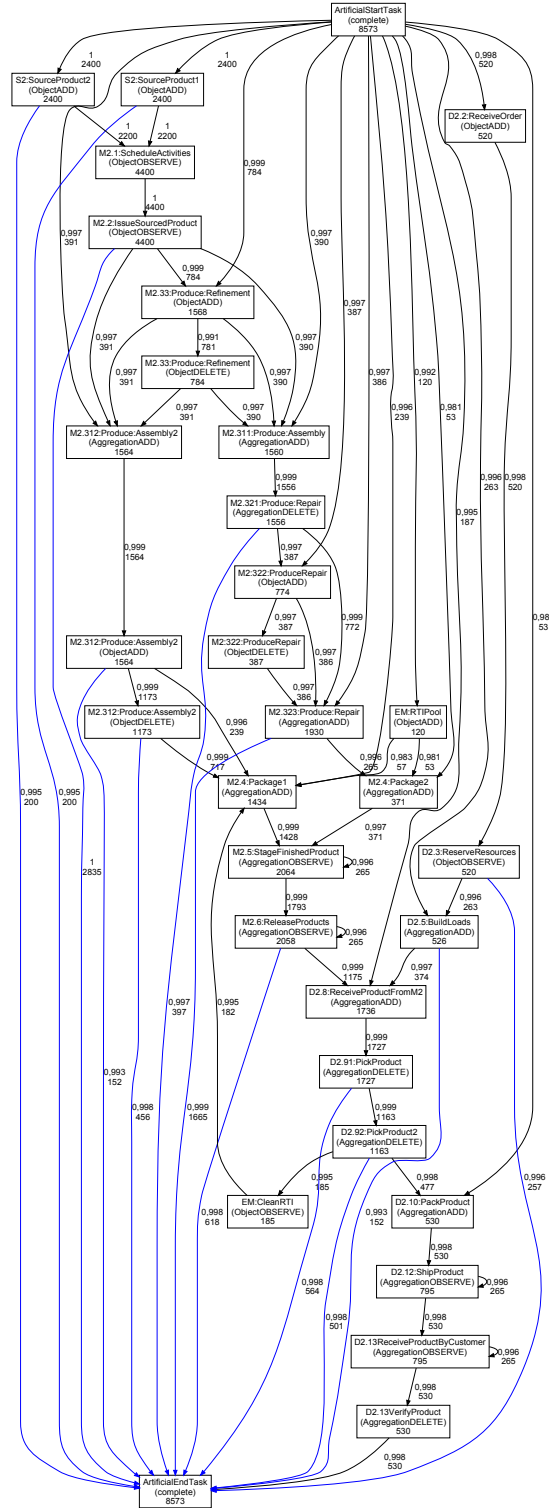


Figure 4.7.: Process model Aut-1 of the overall product flow

4. Process Discovery with Process Mining

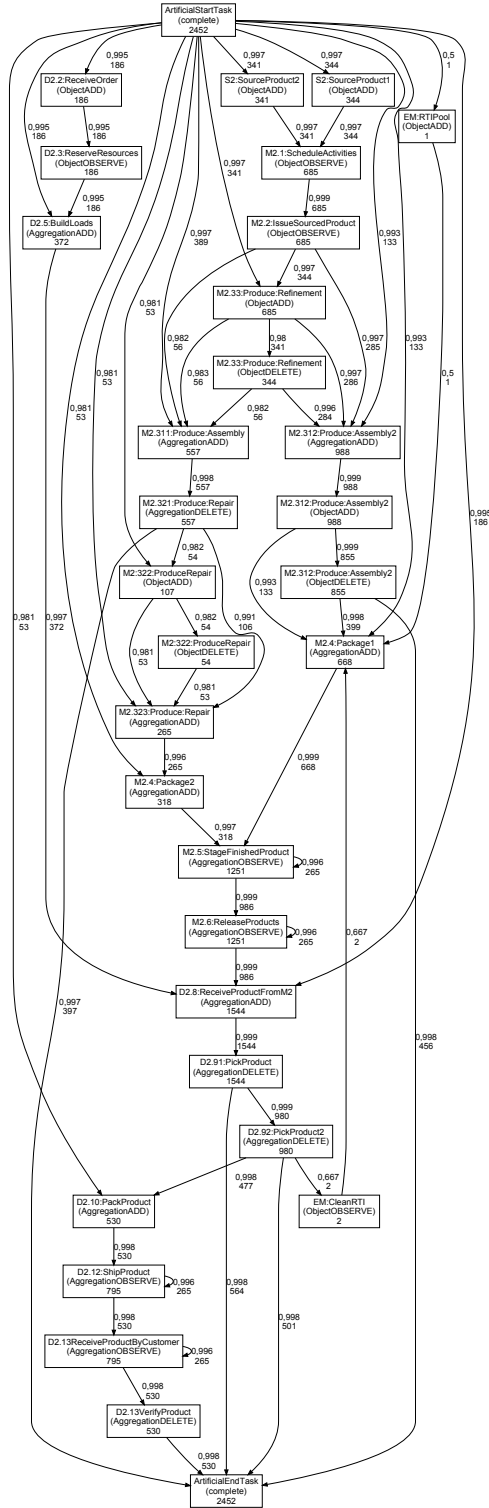
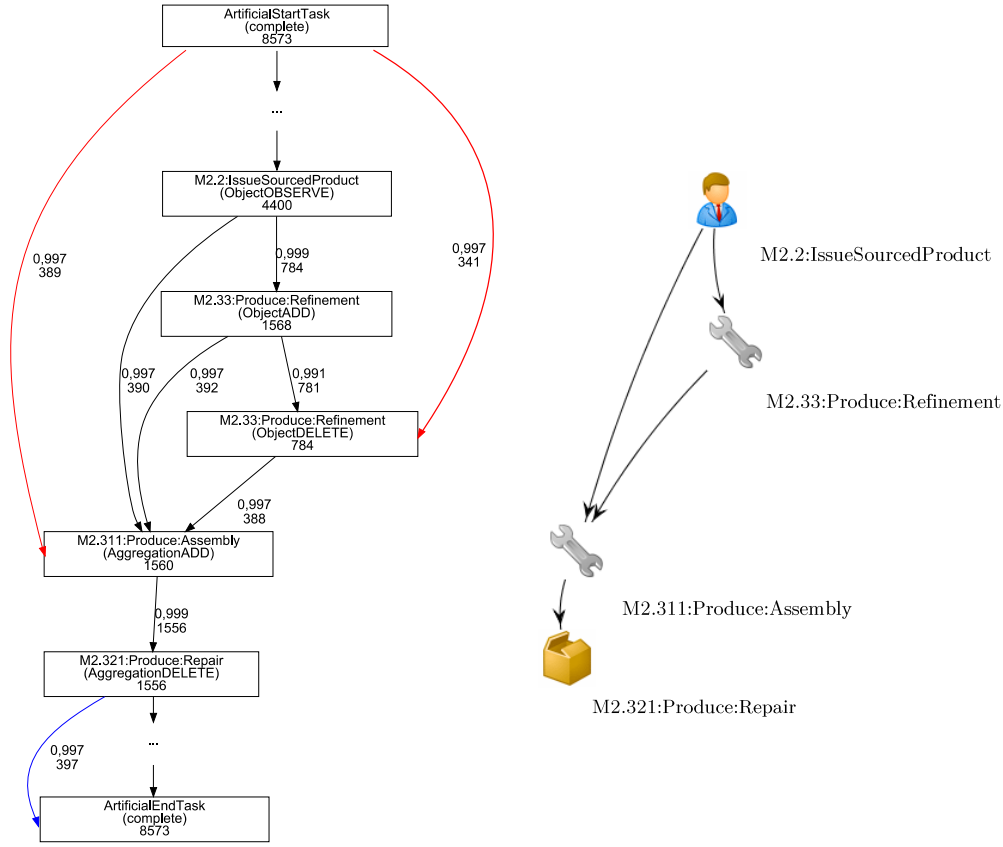


Figure 4.8.: Process model Aut-1 of completed instances

Figure 4.9.: Excerpt of models M'_3 and M_2 of Aut-1

assembled as vehicle doors. In the sample of events, 392 galvanized plates from node $M2.33:Produce:Refinement$ and 390 loudspeakers from node $M2.2:IssueSourcedProduct$ served as assembly input. The transition going from $M2.33:Produce:Refinement$ to $M2.311:Produce:Assembly$ represents the flow of 388 virtual sourceEPCs. These virtual sourceEPCs are necessary to reconstruct the transformation from an unrefined plate to a galvanized one. Every assembly in $M2.311:Produce:Assembly$ has four EPCs: one EPC (i. e., door), two childEPCs (i. e., door handle and galvanized plate), and one sourceEPC (i. e., plate). In the excerpt's graph, the frequency 1,560 in the node indicates that during the simulation 390 doors were assembled in $M2.311:Produce:Assembly$. Since one door handle was missing, two of the 392 galvanized plates remained at assembly. Because of simulated delays in the production process (e. g., capacity constraints), only 389 doors (1,556 electronic product codes) were passed to the node $M2.321:Produce:Repair$. One door remained in node $M2.311:Produce:Assembly$. Patently, the quantities in the model depend on the simulator's settings.

The verification of the reliability of the transitions is now addressed. For example, 389 of 390 doors produced in $M2.311:Produce:Assembly$ arrived at $M2.321:Produce:Repair$. The reliability of .999 shown in Figure 4.9 indicates that the relationship between the

4. Process Discovery with Process Mining

two activities is virtual certain. Since one door remained incomplete, the reliability, though not precisely equal to 1, is extremely near it.

In sum, model M_2 is correctly and completely reproduced in model M'_3 . Model M'_3 certainly manifests additional information; however, these anomalies, though inherent to assembly and disassembly results, do not prevent correct and complete product flow reconstruction. That said, they are needed to reconstruct the virtual flow of sourceEPCs and childEPCs. Obviously, requirements R-1 and R-2 are met.

Requirement R-3: Filtering Events

Here there will be a limit to the EPCIS events that were queried from the EPCIS event repository. Figure 4.11, for example, shows the process model of the first tier supplier responsible for activities M2.1 - M2.6 and D2.1 - D2.10.

In addition to the overall product flow shown in Figure 4.7, Figure 4.12 includes TransactionEvents, among which exemplars are activity M2.2:IssueSourcedProduct (TransactionAdd) and activity M2.312:Produce:Assembly2 (TransactionDelete). The former associates the issue of sourced products with the bill of materials, the latter disassociates the activity from the business transaction production plan.

Requirement R-4: Efficiency of the Algorithm

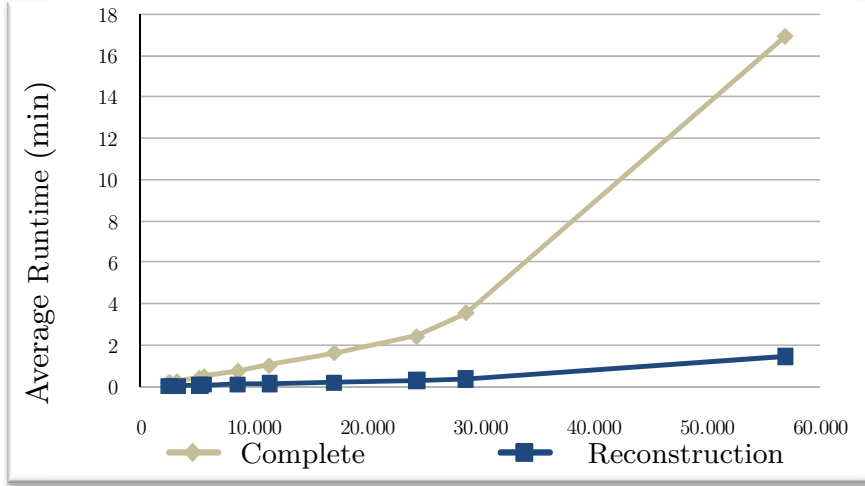
Running the supply chain simulator numerous times produced entries in the EPCIS repository with varying numbers of events. As shown in Figure 4.10, the conversion of algorithm *EPCIS2MXML* took between 15 seconds and 17 minutes depending on the number of events.

At runtime, three phases of computation were performed: the query of the events of the EPCIS event repository, the conversion of the events, and the recording of the MXML file.

Figure 4.10 illustrates that both the runtime of the plug-in (i.e., all three phases) and the share of conversion grow as events grow. True, initially runtime raises little it being not much more than line with the size of events. However, at a size of approximately 30,000 events, growth of runtime powerful takes off. ProMimport's virtual file system explains this change in rate. The file systems swap files, causing the reduction of productivity. As the event size grows, the swap files consume more resources than the conversion. Regardless of this resource allocation, the *EPCIS2MXML* implementation processed the size of EPCIS events simulated on a personal computer (PC).

Requirement R-5: Product Life Cycle Approach

Requirement R-5 mandates the reconstruction of process models that are specific to products, thus maximizing the information degree of model M'_3 . As shown in the upper left-hand corner of Figure 4.6, it is possible to filter products of a product class of a given organization. For example, filtering the process *sgtin-000001.0000004* restricts the overall product flow to that of loudspeakers (product class *0000004*) of the first tier supplier (company prefix *000001*). The product life cycle approach provides a model

Figure 4.10.: Runtime of algorithm *EPCIS2MXML*

hierarchy the range of which spans, at one end from single to combined product flows of one company to, at the other, the overall product flow of a supply chain.

Figure 4.14 (model a) shows the life cycle of returnable items, Figure 4.14 (model b) that of doors, and Figure 4.13 that of loudspeakers. The first model portrays that the returnable items remain at the first tier supplier – though they return to the RTI pool when cleaned – and accompany the other products during production and delivery.

Since the loudspeakers are semi-finished products, their life cycle begins with the sourcing activity *S2*, runs through the production process *M2*, and ends with the delivery process *D2* as part of the end product, namely the car. As loudspeakers are integral components of various types of doors, the process is more complex than that of the doors. For example, both activities *M2.311:Produce:Assembly* and *M2.312:Produce:Assembly2* are involved in processing loudspeakers, whereas only activity *M2.311:Produce:Assembly* is involved in door manufacturing.

Algorithm *EPCIS2MXML* merits careful appraisal. Its output has been compared above to the pre-established result, namely model M_2 . The event log generated by algorithm *EPCIS2MXML* has been taken as output and the process model was derived from that event log. It is noted that both outputs has demonstrated that the conversion of the EPCIS events coincides with the expected results, even when aggregation and disaggregation operations have been involved. Requirements R-1 and R-2 are met.

Since the algorithm provides filter parameters to query only certain events from the EPCIS repository, the algorithm satisfies requirement R-3. The algorithm can be used on a PC and is reasonable in computation time in a reasonable dispatch – requirement R-4 is met. Finally, the product life cycle approach allows the process mining algorithm to reconstruct process models that are specific to products of a given organization – requirement R-5 is met.

4.2. Process Discovery

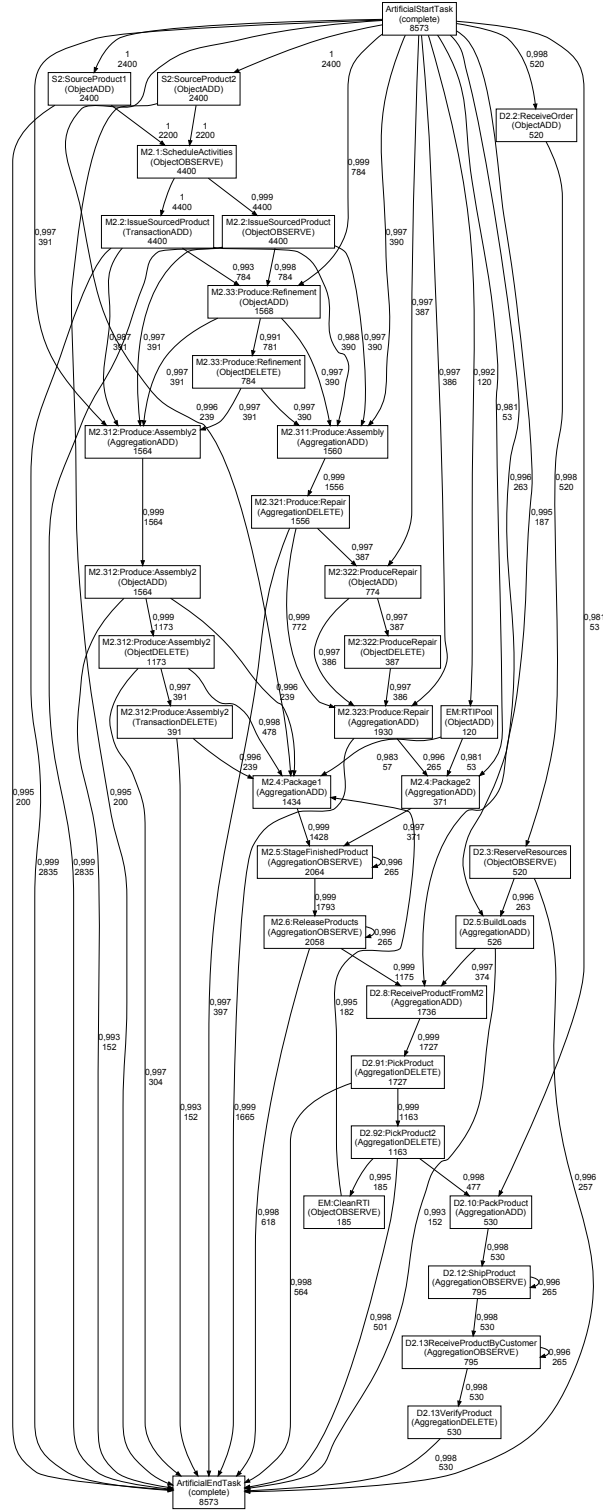


Figure 4.12.: Process model Aut-1 with *TransactionEvents*

4. Process Discovery with Process Mining

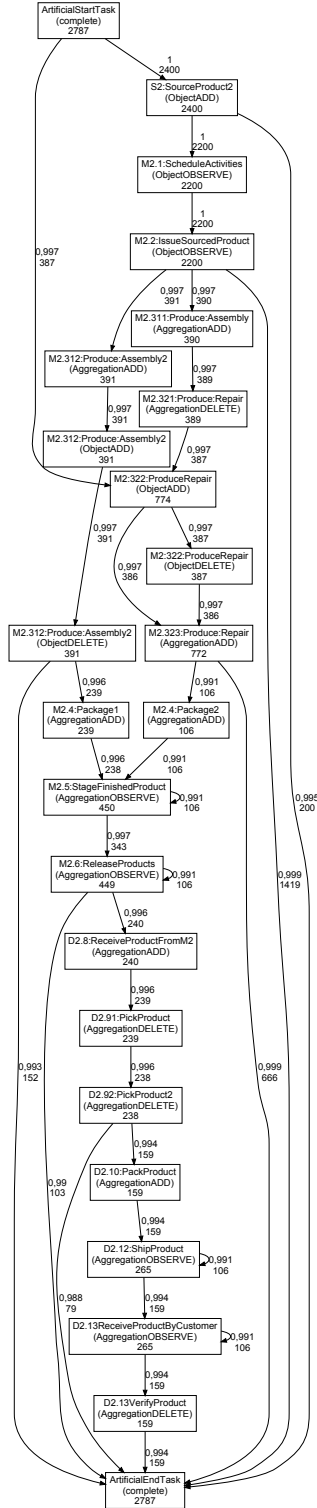


Figure 4.13.: Process model Aut-1 of loudspeakers

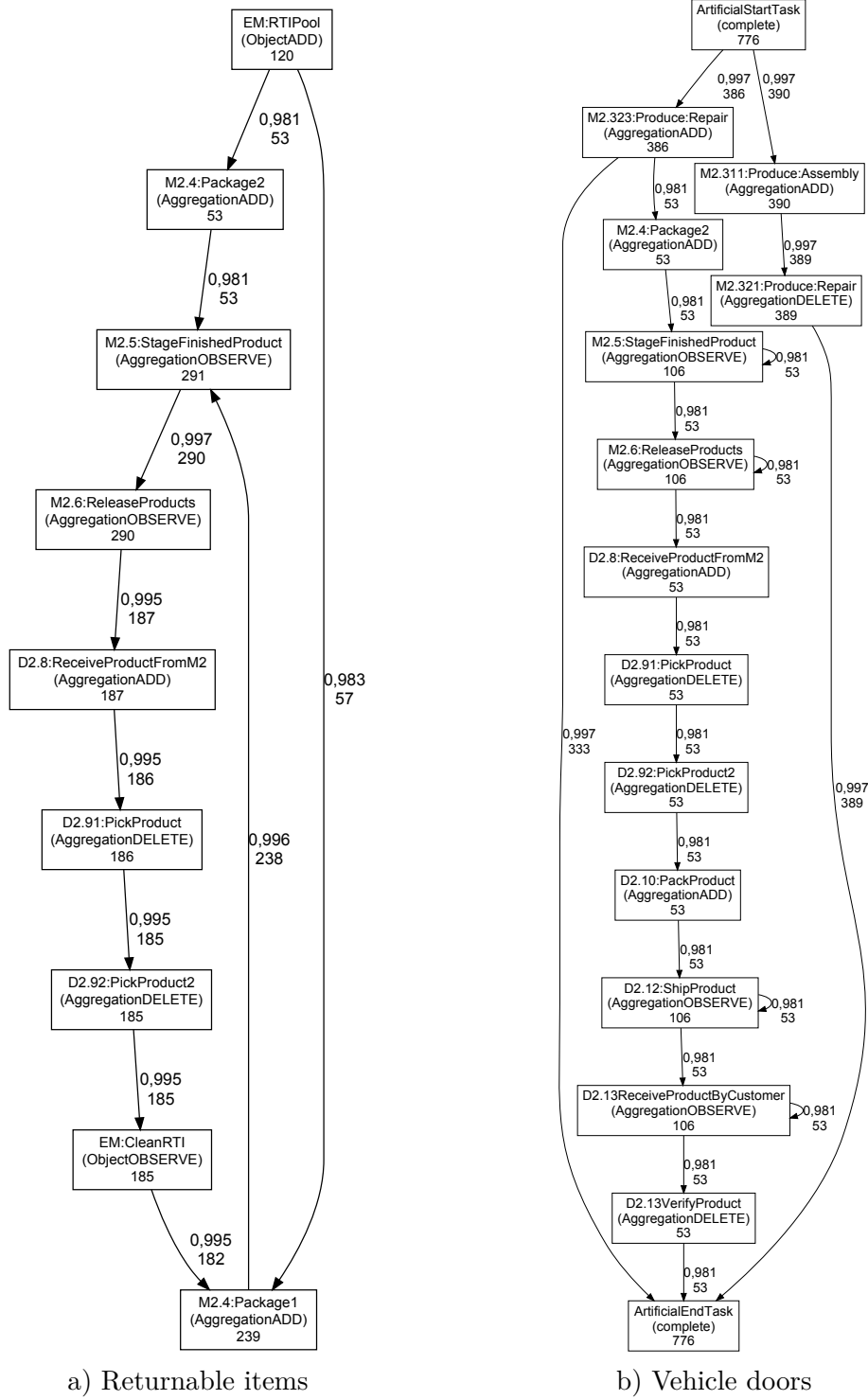


Figure 4.14.: Process model Aut-1 of returnable items and vehicle doors

4. Process Discovery with Process Mining

Evidently, algorithm *EPCIS2MXML* deals EPCIS events, that is, *ObjectEvents* and *AggregationEvents*, very well indeed, especially regarding aggregation, transformation, and disassembly operations. It could be argued that the MXML event field *Workflow-ModelElement* does not lend itself to an appropriate articulation of *TransactionEvents* as they interrupt the product flow. Unfortunately, MXML includes a structure not devoid of business related information to deal with the complexity inherent in a business process. The following section establishes that the algorithm is applicable in an operational environment.

4.2.2. Pharmaceutical Drug Delivery

The data collected during the pilot of use case Pha-1 has been used to derive model M_3 , which reconstructs the product flow of the pharmaceutical supply chain. The to-be supply chain model depicted in Figure 4.2 is henceforth denoted as model M_2 . Appraisal of real-world EPCIS events sheds light both on the expediency of the approach when huge amounts of data⁵⁵ are involved and on the extent to which mining results offer valuable information for process improvement.

The data contained both *ObjectEvents* and *AggregationEvents*, these events representing products that were coded with the EPC types SGTIN and SSCC. The *bizLoc* fields of the events described an organization and its location with a Global Location Number (GLN).

Processing this data meant first cleansing it. This cleansing dealt with missing EPCs and with *AggregationEvents* in which the aggregate identifier was listed within its own *AggregationEvent*. The cleansed data was then loaded into an EPCIS repository installed locally from which it was converted to the MXML format by the *EPCIS2MXML* implementation with granularity setting *coarse-grain*. The lowest level of detail was selected because the reconstruction had to result in a model having the same general level of detail as model M_2 . The coarse-grained conversion also served to protect aspects of confidentiality. The fitness yielded a value of .42. To understand this moderate value, the conversion results are now discussed by comparing model M_3 and model M_2 . The clear structure of model M_2 allows for differences between the models to be ascertained either visually or with a naked eye.

Requirements R-1 and R-2: Correctness and Completeness

Figure 4.15 shows model M_3 of the supply chain of the Pharma Traceability Pilot. The GLN details were replaced by organizations' names and sites, facilitating an enhanced comparison of model M_3 and model M_2 . All quantitative details were removed from model M_3 to preserve confidentiality as to product volumes.

Model M_3 visualizes the read points where products were packed and unpacked. These packing and unpacking activities were represented by *AggregationEvents* with an action Add or action Delete. In addition, products were observed to ensure their traceability. This observation was represented by *ObjectEvents* with action Observe.

⁵⁵ Note that the number of RFID events grossly exceeds that of the other use cases.

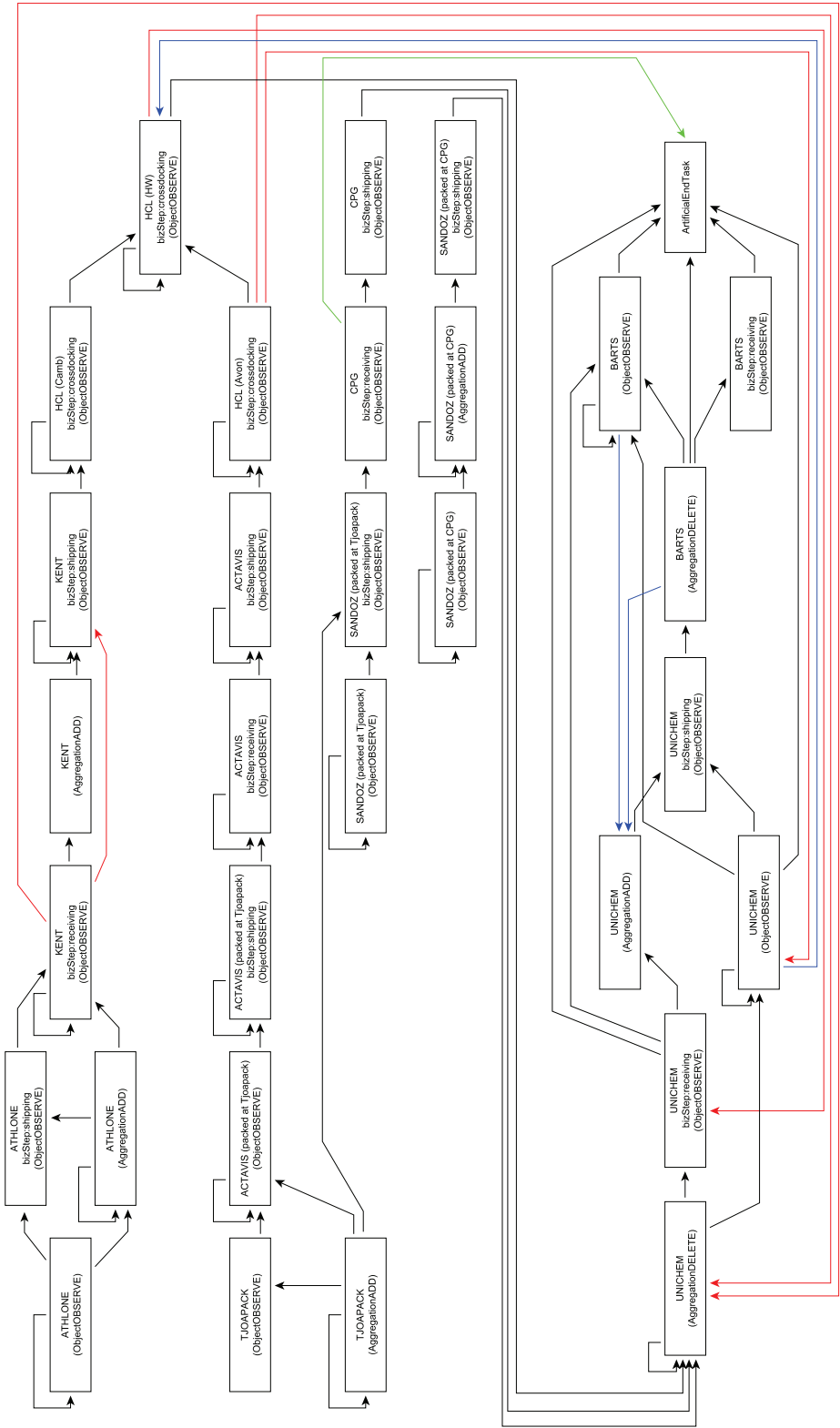


Figure 4.15.: Process model Pha-1 of the pharmaceutical supply chain

4. Process Discovery with Process Mining

One notes a high degree of correspondence when comparing the two models. The individual supply chains of model M_2 are clearly presented in model M_3 :

1. Athlone \rightarrow Kent \rightarrow HCL (Camb) \rightarrow HCL (HW) \rightarrow UniChem \rightarrow Barts
2. Actavis / Tjoapack \rightarrow HCL (Avon) \rightarrow HCL (HW) \rightarrow UniChem \rightarrow Barts
3. Sandoz / Tjoapack \rightarrow CPG \rightarrow UniChem \rightarrow Barts

Ten deviations, which can be classified in four types, were noted:

1. Skipped Activities

Model M_3 contains additional arcs that skip some activities of the process model. These additional arcs indicate product flows that are not described by model M_2 . Figure 4.15 represents these arcs in red.

2. Move Backwards

Additional arcs in model M_3 indicate that products move backwards in the supply chain. These arcs are represented in blue in Figure 4.15.

3. Iterations

Model M_3 consists of a number of activities referring to themselves.

4. Untimely Ending Product Flows

The artificial end event inserted during the reconstruction of model M_3 reveals product flows that ended prematurely. All but one of the arcs that represent these are obviously explicable since the products finally arrived at the hospital (e. g., due to misreading and destruction of transportation packing). Figure 4.15 represents, in green, the seemingly inexplicable deviation.

These deviations warrant explanations – in particular one seeks to learn whether they are reducible to the conversion process or whether the effects can be explained by the data. Each deviation is now analyzed.

Deviation 1 (Skipped Activity)

The activity Kent (AggregationAdd) is skipped. Obviously this deviates from model M_2 since all products from Athlone should have been aggregated at Kent. *

Deviation 2 (Skipped Activity)

An arc goes from activity Kent bizStep:receiving (ObjectObserve) to UniChem. Skipping the logistic service provider (LSP) deviates from model M_2 . *

Deviation 3 (Skipped Activities)

The location HW at the LSP HCL is skipped. Since all products ought to have run through this location, the direct product flow from HCL (Avon) to UniChem deviates from model M_2 . In this case, skipping twice indicates that the product flow skipped two activities. *

Deviation 4 (Skipped Activities)

There are a number of additional arcs indicating skipping, for example, within the activities of the organizations Tjoapack, Athlone, and UniChem. *

During the execution of the pilot a variety of causes for skipping was observed, such as unavailable network connection and unreliable hand scanners (Lilley et al., 2008, p. 49 ff.; Jenkins et al., 2008, p. 60 ff.), both of which prevented the electronic product codes from being recorded. In addition to not being triggered, a number of events were not stored in the repository. These missing events are responsible for additional arcs that skipped activities in model M_3 – they explain deviations one, three, and four.

Arcs between activities within an organization do not generally deviate from model M_2 . These arcs stem from the fact that the coarse-grained setting accounts for the business step, a fact not present as to model M_3 . Deviation two, for example, cannot be explained by reading errors appearing sporadically because four activities were skipped. A scrutiny of the event data showed that the effect on the algorithm *EPCIS2MXML* was caused by faulty data. A number of the pallets shipped from Athlone to UniChem for distribution activity, that is, UniChem (AggregationDelete), were recorded with the wrong time stamp. Therefore, the corresponding event was ordered between the activities Kent bizStep:receiving (ObjectObserve) and Kent (AggregationAdd). Algorithm *EPCIS2MXML* broke the connection between the pallet and the products on it, this in response to the disaggregation event. As a result of this disaggregation, all subsequent events related to the pallet were unconnected with the products that had been on the pallet. Though the pallet ran through all specified activities, the corresponding cases that had been on the pallet could not be traced until they arrived at UniChem. The above vividly illustrates the potential disruption when the wrong time stamps are used.

Deviation 5 (Move Backwards)

Products seemingly were shipped from UniChem back to the LSP HCL. *

Deviation 6 (Move Backwards)

Products seemingly were moved from Barts back to UniChem. *

Neither of these two deviations is covered by model M_2 . Like deviation two, deviations five and six are caused by faulty time stamps. A product may appear to have moved backwards because, for example, a faulty time stamp at point B incorrectly indicates that it was there *after* instead of before it moved on to point C – if this occurs, and if further observations are dated accurately at other locations, the algorithm derives a move backwards from this flawed chronology.

Deviation 7 (Iterations)

There are iterations at activities of type AggregationAdd at Athlone, Tjoapack, and Sandoz. *

Deviation 8 (Iterations)

An iteration exists at the activity UniChem (AggregationDelete). *

Deviation 9 (Iterations)

Iterations exist at activities of type ObjectObserve at most of the organizations in the supply chain. *

As with deviation four, the low level of detail of model M_3 is the basis for the observed iterations. Different effects of the coarse-grained conversion on the reconstruction of model M_3 exist: multistage packing and scanning on different levels of product packing. In general, products are aggregated (i.e., items, cases, and pallets) on all levels of product packaging (Jenkins et al., 2007c, p. 20) in the supply chain. The coarse-grained conversion also results in a generalization about the EPC types – such multistage packing and unpacking events are reconstructed as iterations at aggregation and disaggregation events, a fact exemplified in deviations seven and eight.

Actions carried out to ensure the traceability of the products, namely multiple product scanning (Jenkins et al., 2008, Section 4.2.3.14) and single case scanning (instead of the pallet label), explain deviation nine (Jenkins et al., 2008, Section 4.2.3.6). These actions resulted in numerous ObjectEvents with an action Add for the products. A more differentiated recording by use of AggregationEvents for pallets would have led to a more exact reconstruction result.

Deviation 10 (Untimely Ending Product Flows)

An arc goes from activity CPG bizStep:receiving (ObjectObserve) to the artificial end event. The arc shows that some products remained at one of the distributors were taken out of the supply chain. *

The pilot evaluation showed that a distributor sold some BRIDGE stock to a third party utterly outside of the pilot (Jenkins et al., 2008, Section 4.2.3.7). Since the flow of these strayed products naturally terminated at the very distributor from which they were physically lost, they vanished from the traceability system. Therefore this misdirection may explain deviation ten.

All deviations alternatively were accounted for by the data – thus, the possibility of conversion errors is excluded.

Requirement R-3: Filtering Events

Filters allow for access to certain events. It is possible, for example, to filter for a specific organization. Figure 4.16 shows on the left-hand side (model a) the product flow of antibiotics manufactured by Athlone.

Requirement R-4: Efficiency of the Algorithm

The overall conversion time of algorithm *BO2MXML* was 28 minutes. The resulting MXML file contained 363,688 events distributed among 24 processes – this is approximately 5.9 times the size of the original EPCIS events. Nevertheless, this vast data can be processed by a PC.

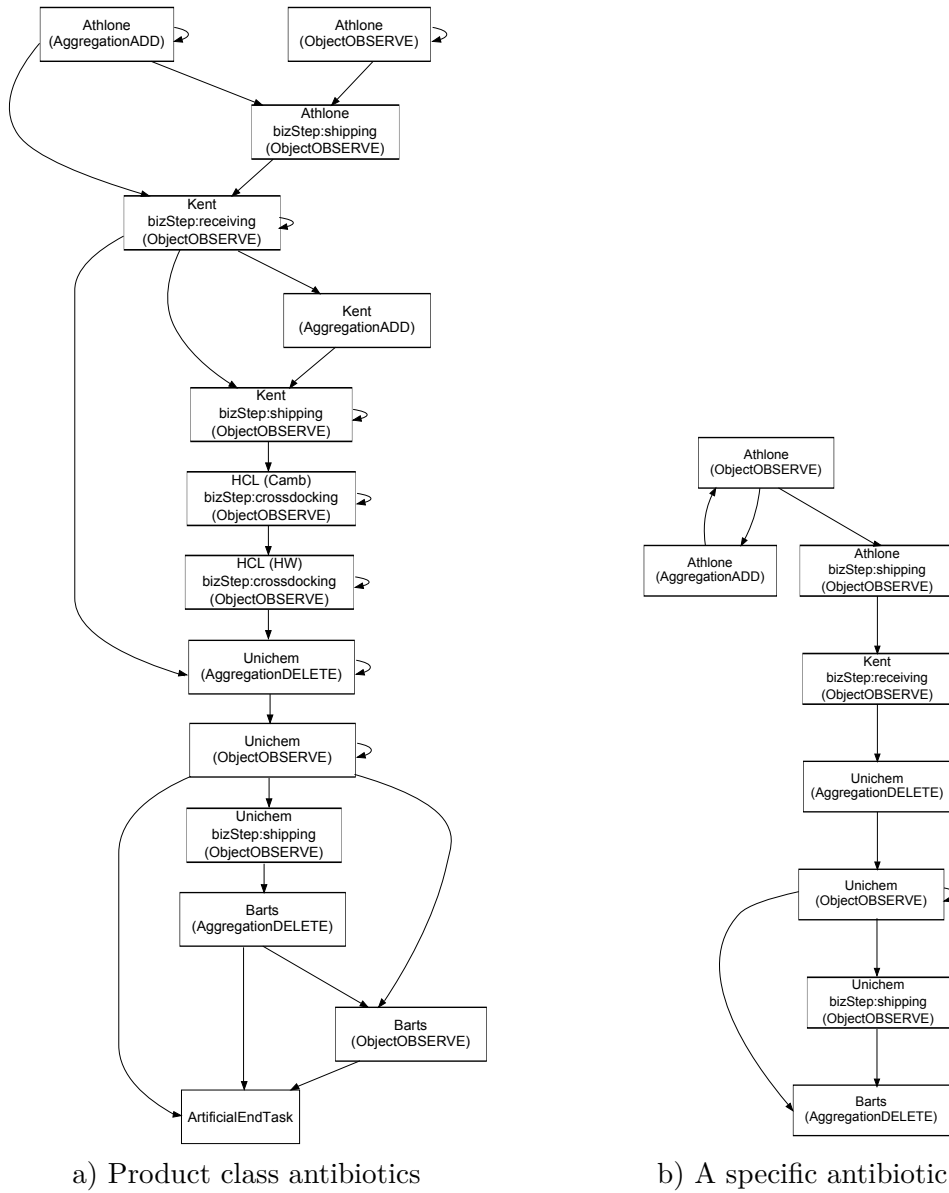


Figure 4.16.: Process model Pha-1 of manufacturer Athlone

Requirement R-5: Product Life Cycle Approach

Figure 4.16 illustrates on the right-hand side (model b) the product flow of a type of Athlone antibiotic ⁵⁶.

⁵⁶ Details on single products and quantities are withheld due to nondisclosure agreements.

Patently, the employment of algorithm *EPCIS2MXML* in an operational environment confirms use case Aut-1's results. The approach's expediency is also confirmed when huge amounts of data are involved, and valuable information for process improvement is provided.

4.2.3. Complaint Handling

Since the IC does not support logging functions, all information pertaining to complaint cases needs to be both identified in the CRM system and prepared for process mining. Algorithm *BO2MXML* has therefore been employed in the SAP CRM system of use case Air-1. The validation adopts an acceptance sampling approach, that is, in industry a common quality control approach and is recommended by quality tools such as Six Sigma. Comparing the sample cases with the document flow recorded in the system determines whether the outcome of algorithm *BO2MXML* yields the expected result.

Requirements R-1 and R-2: Correctness and Completeness

The conversion of algorithm *BO2MXML* to the MXML format is the same as that of algorithm *EPCIS2MXML* and therefore needs not be proven again – one needs only to verify the concept of reconstructing the document flow.

A key aspect of acceptance sampling is that it uses statistical sampling to determine whether to accept or reject a production lot of material. Sampling involves the inspection of a small number of units to make decisions about the acceptability of a larger number. Sampling facilitates appraisal of algorithm *BO2MXML*'s correctness and completeness.

The sampling cases have to be carefully selected to be confident that they are indeed representative. The ProM export plug-in *Group MXML Log* groups sequences expressing the same sequence of activities. For instance, if an event log contains the three sequences (a, b, c, b) , (a, b, c, c) , and (a, b, c, b) , then the two sequences (a, b, c, b) and (a, b, c, b) are grouped.

Grouping cases reduced the cases from 4,650 to 397 activities, while maintaining the same variance inherent in the event log. Grouping was advantageous because same sequences were barred from the sample. Approximately five percent of the 397 cases, that is, 20 cases, were randomly chosen for manual inspection – naturally, each case had the same chance to become a sample element. The inspection of the sample cases determined the number of cases that did not conform to the document flow in the CRM system. A case was accepted as true only when positive observations were made – thus meaning that the document flow was correctly and completely reconstructed.

The document flow as depicted in Table 3.5 is now analyzed. The *Log Inspector* available in ProM allows a single case to be browsed. Figure 4.17 shows the example of the first case, that is, $\pi = \{137, 138, 139, 141, Ross\}$. Activity *Create activity Cust. Relations* corresponds to line four of the table, activity *Identify account* to that of line three, activity *Create activity Cust. Relations* to that of lines six and ten, activity *Edit activity Cust. Relations* to that of lines two and seven, activity *Create activity Cust. Relations* to that of lines nine and eleven, activity *Edit activity Cust. Relations*

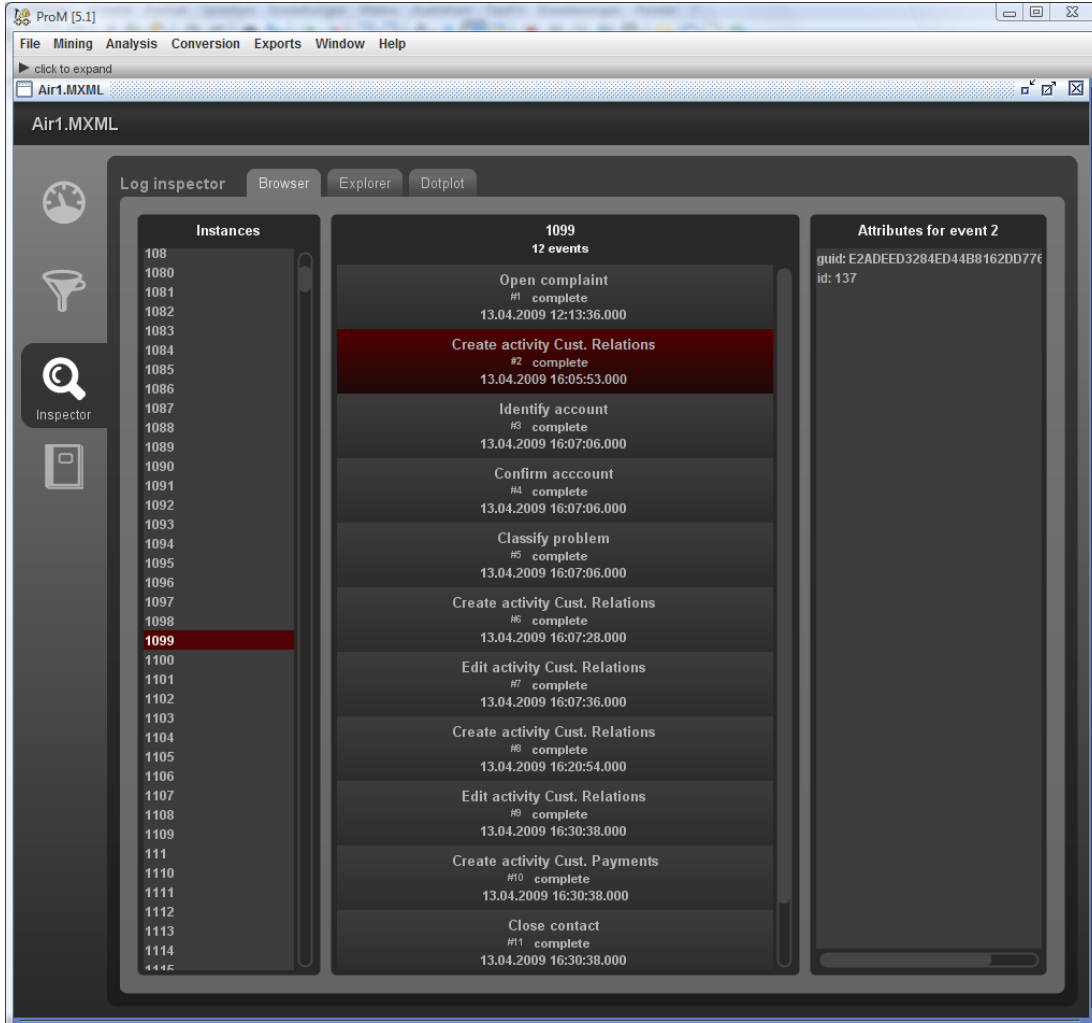


Figure 4.17.: Log Inspector plug-in of model Air-1 in ProM

to that of lines one and eight, and activity *Create activity Cust. Payments* to that of lines five and fourteen. The document type *CrmaAnchorObject* indicates that activity Cust. Relations 140 belongs to a different case, one with activity Cust. Relations 142 as successor. The remaining activities depicted in Figure 4.17 do not have a link to the document flow because they result from the reconstruction of the activities. Evidently, the relations are completely and correctly reconstructed, as a result of which this case is marked as true.

Each sample case was investigated in this way. Algorithm *BO2MXML* did not produce an erroneous result in any of the samples. Because of this, the partition approach is considered to be a good representation for the reconstruction of the interrelated chain of IC elements.

4. Process Discovery with Process Mining

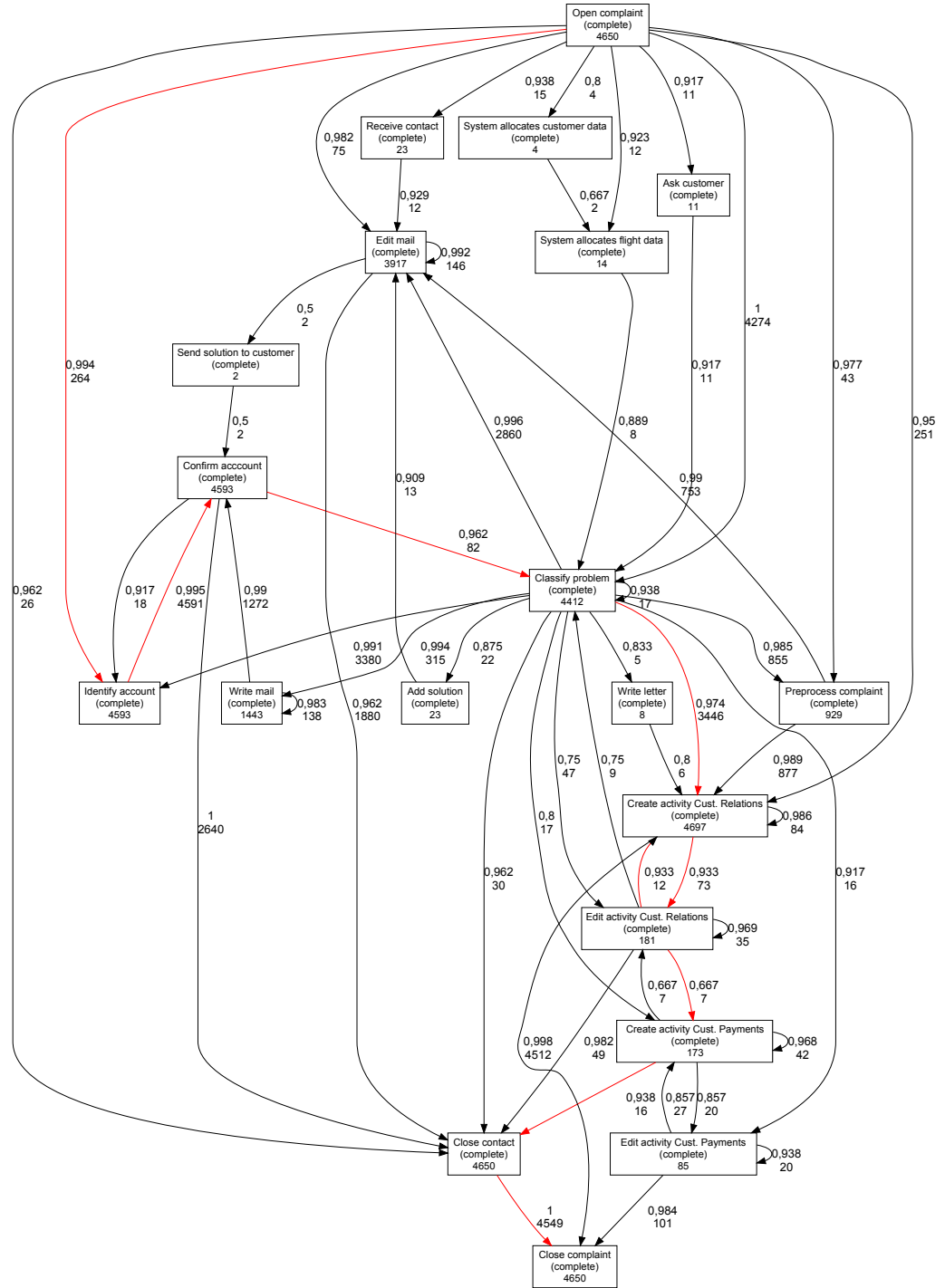


Figure 4.18.: Process model Air-1 of the complaint handling

Model M_3 shown in Figure 4.18 exhibits the behavior from the example in the above

discussion – it is highlighted in red. Note that one discovers that there were activities with identical time stamps that were represented in the same manner as when they were in a sequential order. Designate the activities *Classify problem* as “A” and *Preprocess complaint* as “B”. The arc from activity A to activity B signifies that the former is the cause to the latter. The *Heuristics miner* reconstructs this causal dependency – algorithm *BO2MXML* infers the activities in such a way that activity A is followed by activity B and activity A never occurs after activity B, that is, $B \rightarrow_L A$, though the ordering relation might be $A \rightarrow_L B$ or even $A \parallel_L B$. Bear in mind that for the purpose of preserving chronological order the preprocessing of data imposes some order on extracted activities.

The fitness of the process model yielded a value of .37. Its moderate value is due to unstructured data and a high degree of freedom on the part of IC agents executing their tasks.

Requirements R-3: Filtering Events

Figure 3.5 illustrates how data is selected by filtering parameters, among which are the CRM activity, role in document flow, period of time, and responsible group. The process model in Figure 4.19 shows the complaint processing of department *Cust. Payments* (i. e., CRM activity type = Cust. Payments).

Requirement R-4: Efficiency of the Algorithm

The conversion of algorithm *BO2MXML* took between seven and fifteen minutes due to varying system capacity of the productive system. Since it is a productive SAP CRM one, additional run time measurements were not feasible. As the *BO2MXML* implementation processed 44,006 activities, it is reasonable to deem that requirement R-4 is met.

Requirements R-5: Representation of Reference Activities

It is troubling that a number of relevant facts were not available because CRM system’s transactions did not have attributes to represent them. Figure 4.18 shows the successful decomposition of activities from the business transaction CRM activities. Note that the ordering of the activities was an approximation because of missing details about the time. To determine if this approximate solution provides information increase, enterprise data based on transactional data is selected. Figure 4.20 depicts the process model without the decomposed activities *Ask customer*, *Confirm account*, *Add solution*, *Classify problem*, *Preprocess complaint*, and *Close contact* all of which belong to the business transaction CRM activity. This omission reveals an important fact: Dependencies cease that existed between the omitted activities *and* (to an extent) between the remaining ones. Take the relationships between *Edit activity Cust. Relations* and *Create activity Cust. Payments*. They were concealed on the transaction level because they were derived by inspection of the cases in their entirety. Having found evidence that the derived information adds value for the process improvement, requirement R-5 is partially satisfied.

4. Process Discovery with Process Mining

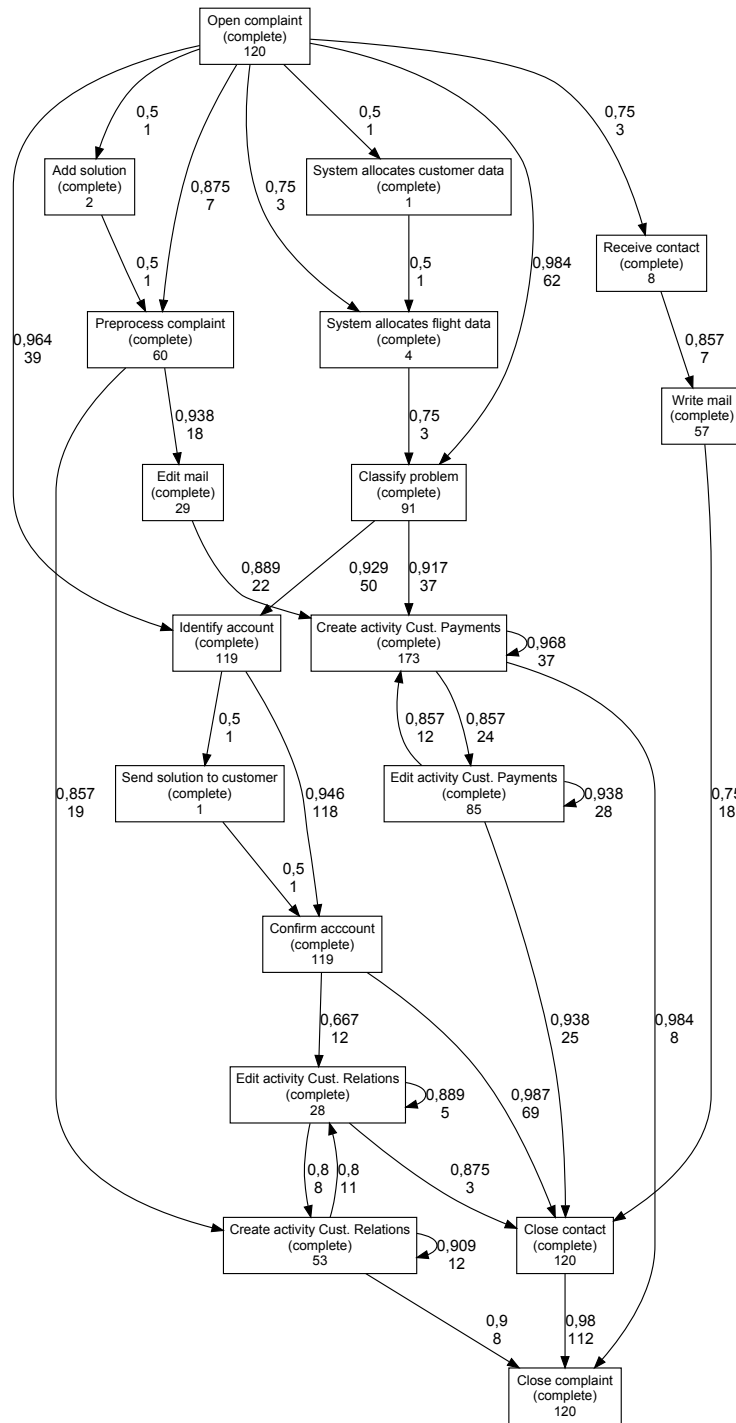


Figure 4.19.: Process model Air-1 of group Cust. Payments

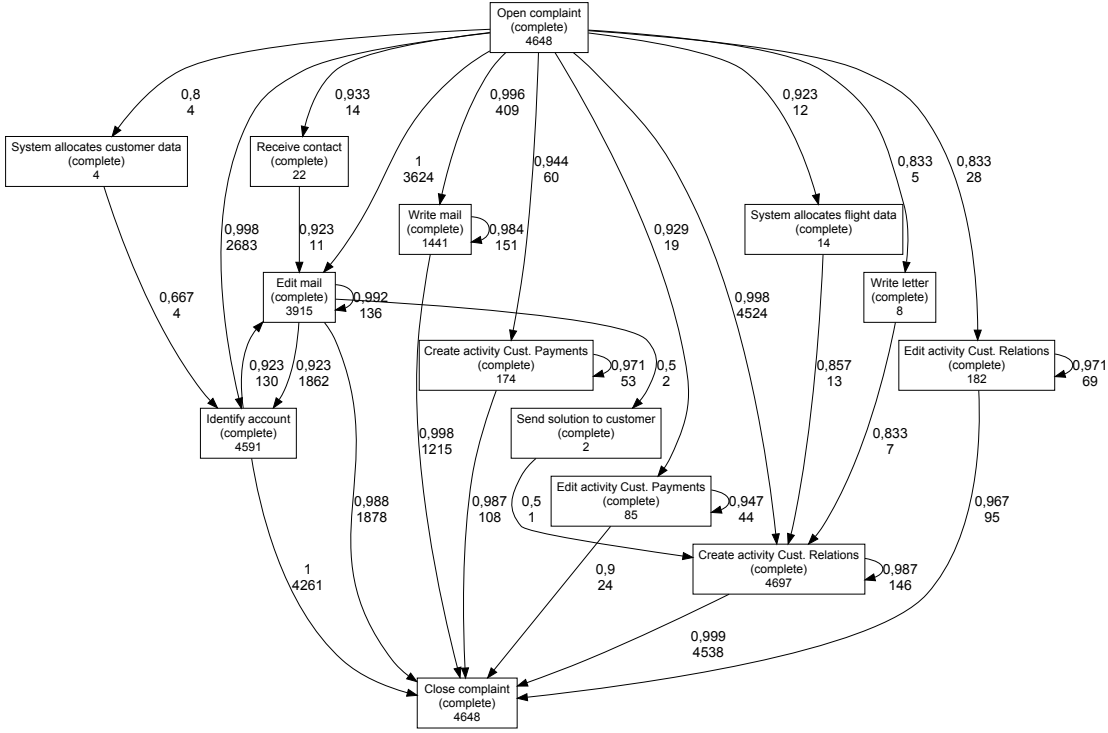


Figure 4.20.: Process model Air-1 on transaction level

Algorithm *BO2MXML* now warrants overall appraisal. Sampling itself verified this algorithm's output's correctness and completeness. Each sample was accepted as true, meaning requirements R-1 and R-2 are met, albeit with the proviso that the above-noted ordering problem exists. Since the algorithm provided filter parameters restricting selection to only certain CRM activities from the IC, the algorithm obviously meets requirement R-3. Algorithm *BO2MXML* is reasonable in computation time, fulfilling requirement R-4. Finally, the successful decomposition of activities from transactions showed that requirement R-5 is somewhat met; the representation of the decomposed activities provided valuable information, granted that only an approximate ordering of the activities could be reconstructed. Algorithm *BO2MXML* therefore deals with enterprise data properly even where compositions and associations are involved.

4.2.4. Service Operation

Taking for granted that the concept of process mining has been theoretically proven (Lang, 2008; Goedertier, 2008; Alves de Medeiros, 2006), this section shows how organizations might improve their processes from process mining using an approved data source, namely workflows.

4. Process Discovery with Process Mining

Service e-mail (henceforth use case “Tel-1”), herein exemplifying all services of use cases Tel*⁵⁷, is described in detail. From the history of action, incidents were selected that were closed within a given period of time, and they were stored in a comma-separated values (CSV) file. It was converted to the MXML format, introduced in Section 3.2.1, by an import plug-in of ProMimport.

Table 4.6 shows the MXML representation of the first line of the history of action (see Table 4.5). The type of incident T1596654, process, and priority are listed in the *Data* element, respectively. The description of the processing step corresponds to the *WorkflowModelElement*. The support group was mapped to the *Originator*, the status was assigned to the *EventType*, and the time and date to the *Timestamp* element.

Table 4.6.: Example MXML event derived from the history of action

```
<AuditTrailEntry>
  <Data>
    <Attribute name="Process type">Failure</Attribute>
    <Attribute name="Process">Retrieve e-mail</Attribute>
    <Attribute name="Priority">Low</Attribute>
  </Data>
  <WorkflowModelElement>1st Level Support</WorkflowModelElement>
  <EventType unknowntype="forwarded">unknown</EventType>
  <Timestamp>2009-10-06T09:47:43.000+00:00 </Timestamp>
  <Originator>Service desk</Originator>
</AuditTrailEntry>
```

Figure 4.21 shows the mined model M_3 .

Model M_3 yielded a fitness of .45 indicating a moderate alignment between the process model and the event log. This value may be attributed to the fact that there was no discernible navigation pattern as to the incident management process, an absence caused by the fine-grained level of detail in the history of action.

To reveal this pattern the underlying structure of the process was analyzed. The scheme capturing all possible actions of the WFMS is called transaction scheme. It was derived from the history of action. Except for the event field *WorkflowModelElement*, each attribute was mapped as done above. Note that the attribute *status* of the history of action was mapped to the event field *WorkflowModelElement*. The resulting scheme revealed the processing of service e-mail. It is shown in Figure 4.22⁵⁸, which, for clarity sake, does not depict frequencies.

When an incident is created, its action is either *in progress* or *forwarded*. In the case in which the ticket has been initiated from one technical interface of the event management, its processing starts with the action *forwarded* – after forwarding, the incident is assigned to a support group. The assignment is documented in the history of action as *assigned*.

⁵⁷ The asterisk denotes all use case numbers of the ISP, that is, Tel-1, Tel-2, Tel-3, and Tel-4.

⁵⁸ See Appendix C for the corresponding Heuristics net.

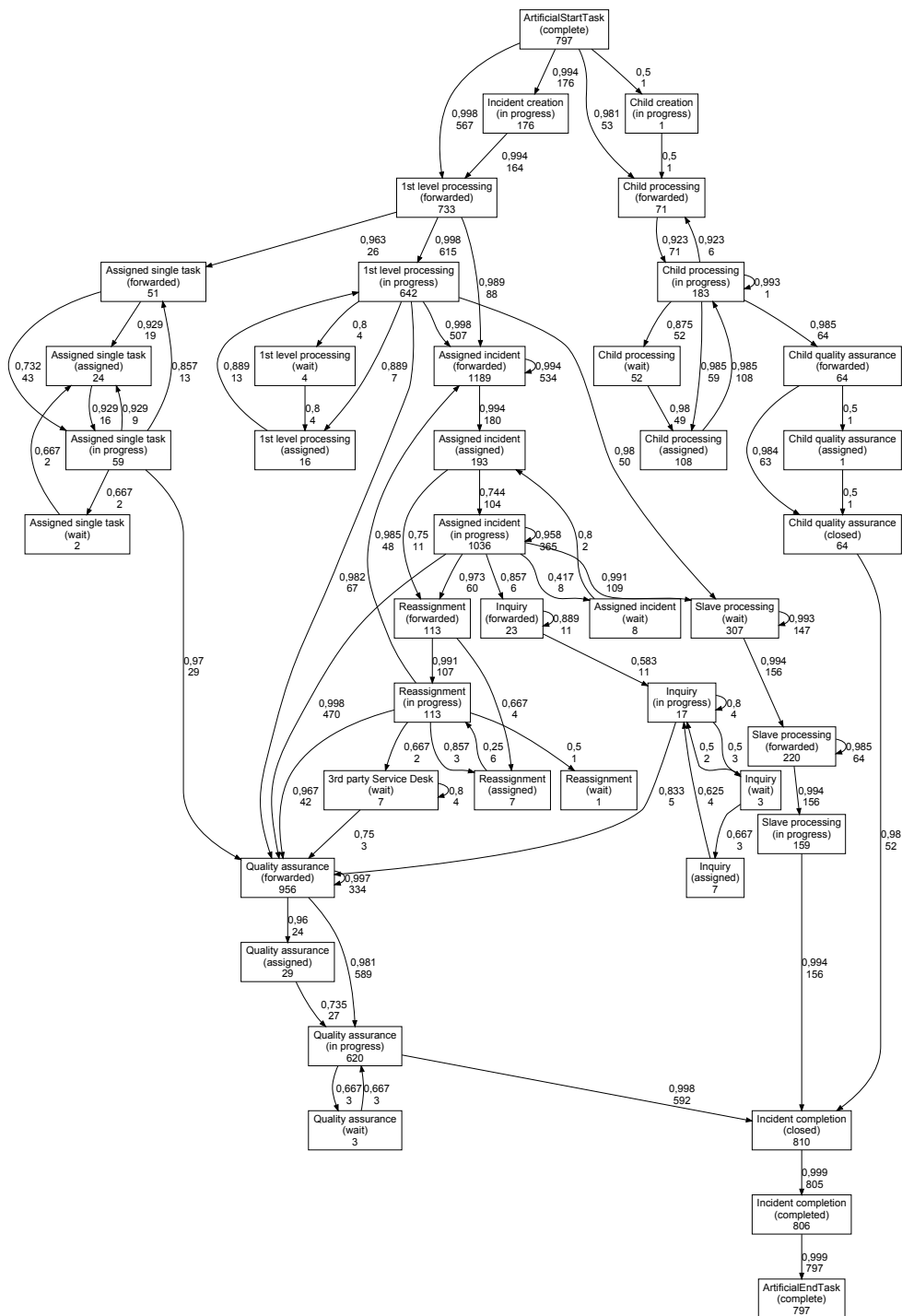


Figure 4.21.: Process model Tel-1 of service e-mail on action level

4. Process Discovery with Process Mining

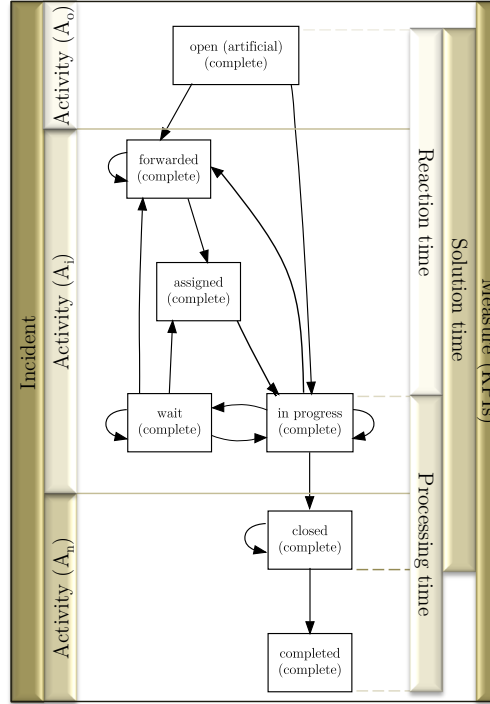


Figure 4.22.: Service operation of an ISP

In the case in which the ticket is not so initiated, it creates an action *in progress*. During – indeed in progress – an incident’s remedy, tickets can become slave tickets. Consider a typical scenario, that in which there is a single incident and many users call and open a ticket. One of the tickets becomes the master ticket, and all others are slave tickets that wait (i.e., action *wait*) until each is automatically closed as the master ticket is closed. Later, the incident itself, once solved, is closed, an action leading to the action entry *closed*. After a final quality assurance the support group completes the incident – this is the action *completed*. Regardless of the actions taken, the WFMS records them according to all relevant information so that a full history of action is maintained.

Figure 4.22’s process model portrays the relations between actions and activities – see the far left-hand. The start action, that is activity A_0 , initiates the proceedings of the incident. The action types *in progress*, *forward*, *assigned*, and *wait* follow the start action and together belong to one activity, denoted as A_i . An ideal incident management process goes through each action but only once. Naturally, in practice repetitions of various action types of the activities were found. For example, one can reassign an activity to a different support group (*in progress* to *forward*) or one can suspend and then resume an incident (*in progress* to *wait*), say, two or three times during the ticket flow. The actions *closed* and *completed* belong to the final activity expressed with A_n . Take the following example: The activity *Incident Creation (in progress)* is followed by the activity *1st Level Support*. The latter activity consists of the processing steps

forwarded, *assigned*, and finally, again *in progress* before a new activity begins.

The history of action has another advantage – one can judge the incident management process’s efficiency and effectiveness and that of its operation (Kütz, 2009). Figure 4.22’s far right depicts the relationship between the history of action and quality measures. Points in time from one action to the next are used to measure various key performance indicators. For example, the solution time is the period between the opening and solving of an incident, which is derived by the actions *open* and *closed*. The processing time equals the solution time minus the reaction time plus the time spent for quality assurance and closure actions, closure action being from *closed* to *completed*.

As seen, the transaction scheme revealed the navigation pattern followed by each case – now this pattern is removed from the very fine-grained model M_3 . A combination of filtering techniques to subsume the fine-grained actions to the granularity level of ITIL activities has been applied. The filters in question permit the comparison of the history of action with ITIL activities. Figure 4.23 illustrates the effect of these techniques, which are available in ProM.

The line at the bottom shows a fragment of the parent-child relationship *on action level*. This relationship is a group of change requests, namely children, linked to one parent change request. Each child has to be processed to close the parent change request. The second line from the bottom shows the child processing after the *Repetitions-to-Activity Filter* has been applied. This filter replaces each direct repetition of the same action, for example, *Child processing (in progress)*, *Child processing (wait)*, and *Child processing (assigned)* by two events, one starting with the time stamp from the first occurrence and a complete one with the time stamp of the last occurrence in the sequence of repetitions. If there is only one occurrence of one type of action (e. g., *Child creation (complete)*), it is replaced by a start and complete event with the same time stamp.

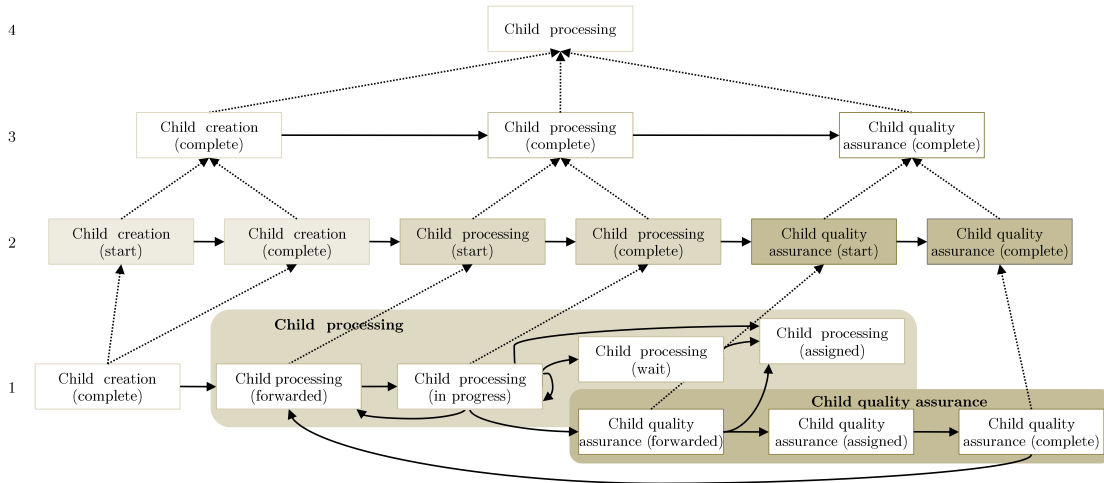


Figure 4.23.: Differences between actions and activities

Due to the Repetitions-to-Activity Filter's work, the number of cases decreased from 797 to 176, signifying a more general process flow. Next, the start events have been discarded through the simple filter (see the third line from the bottom) – which leads to the activities (depicted in the top line of the figure) each of which is atomic. Note that model M_3 on *activity level* is denoted as M'_3 (see Figure 4.24). Applying the filtering techniques resulted in a fitness of .49, in contrast to the previous level of .45. This very slight increase indicates that the abstraction from the ticket flow does not substantially alter the process model's quality.

Figure 4.25 juxtaposes model M_3 with M'_3 . The left-hand side shows the fine-grained model M_3 on the action level and the right-hand side shows the coarse-grained model M'_3 on activity level.

The systematic analysis as described in this section was also carried out for the services S_1 , S_2 , and S_3 , the salient results of which follow. It was possible to extract process models both on action level and activity level for all three services. The properties of the services, such as number of events, selected period, are listed in Table 4.2. As is indicated in it, applying the filtering techniques reduced the number of events, that is, M'_{3-S_1} 8,071 events, M'_{3-S_2} 19,386 events, and M'_{3-S_3} 24,081 events. The fitness of the models both on activity level – .33 for S_1 , .32 for S_2 , and .31 for S_3 – and on action level – .47, .49, and .28 respectively – yielded values strongly resembling those of service e-mail.

4.3. Conclusion

Five case studies have been conducted at diverse organizations – these studies will facilitate interested parties' appraisal of process mining's potential, especially with respect to its ability to exploit new data sources and to optimize extant sources.

These studies have proven the expediency of process mining regardless of sector, data source, process type, product, and services. Process modeling with process mining gives an accurate account of the as-is condition in the organization – process mining may serve as a basis for improvement measures and their corresponding post-control. Process mining for continual process improvement is most welcome because the technique both is objective and can be used in a highly automated way. These two features provide reproducible and repeatable results, which are, after all, essential for keeping process models up-to-date.

Use cases Tel* show that workflows are ideally suited to process mining. However, flexible or complex processes often result in “spaghetti-like” process models that are hard to fathom, and with that prevent a comprehensive use for process analysis. To obtain meaningful models particularly in the context of process improvement, the focus has been on generalization from less important information. Uncommonly used (i.e., the status is the *WorkflowModelElement*), process mining has enabled the identification of the transaction scheme of the incident handling. Filtering techniques have assisted in generalizing from this scheme suffering no observable loss of quality. In this way, the fine-grained actions have been compressed to meaningful activities the granularity of

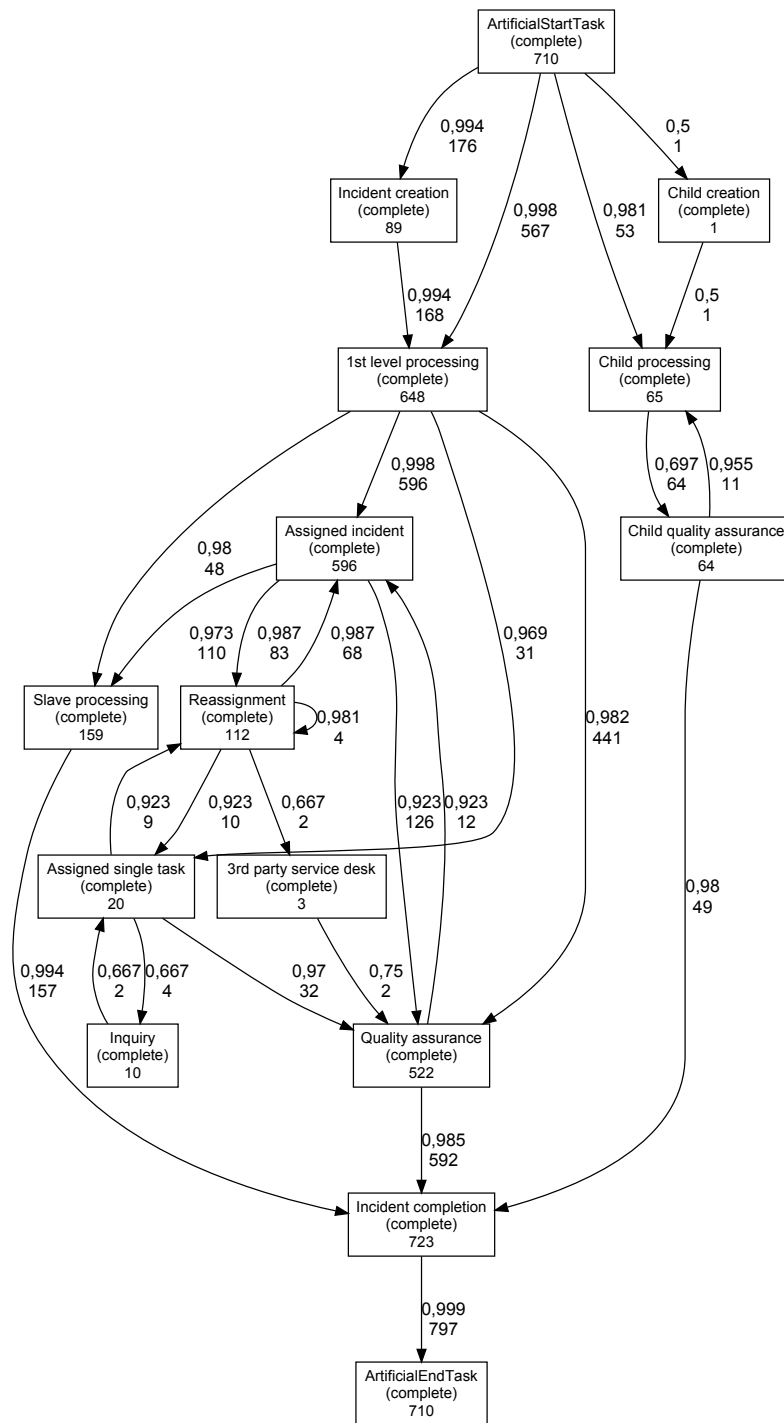


Figure 4.24.: Process model Tel-1 of service e-mail on activity level

4. Process Discovery with Process Mining

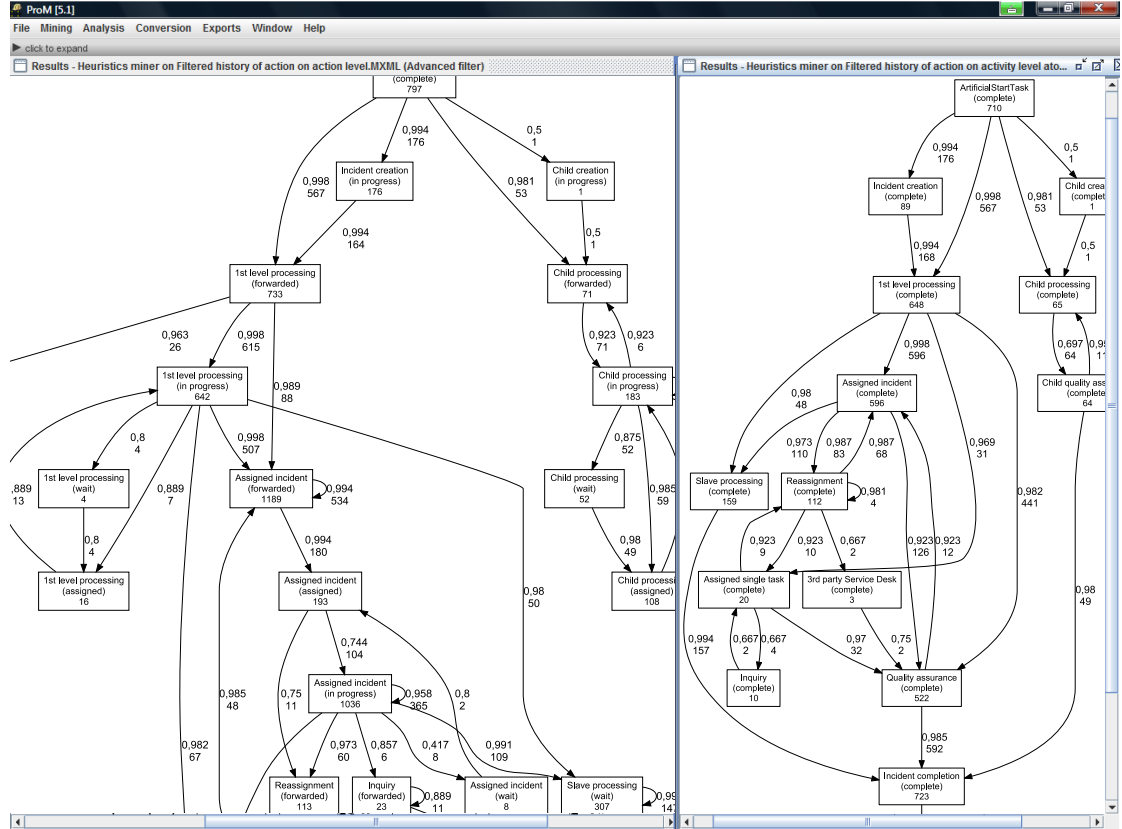


Figure 4.25.: Process models Tel-1 on different granularity levels

which can be compared to those of a reference model. Two types of process models have been consequently derived that provide new possibilities for process improvement: models on action level and models on activity level.

Despite all benefits, there are also caveats. The following two reared their heads as to process mining implementation: data preparation effort and process mining's purview on IT-enabled processes.

The first obvious drawback is that the results of process mining strongly depend on data quality. Data preparation is both strenuous and very time-consuming. Data preparation requires identification of activities and cases. And the data preparation⁵⁹ for the use cases studied for this dissertation took, depending on the particular case, from between 40 % and 80 % of the total time devoted – though one hastens to add that these times, bad as they seem, do not greatly exceed those of data preparation as to data mining projects. It is well-settled that data preparation addresses an important issue in the area of data mining. The Cross Industry Standard Process for Data Mining (2008,

⁵⁹ According to the CRISP-DM model, the activities select, cleanse, construct, and format data belong to the data preparation stage.

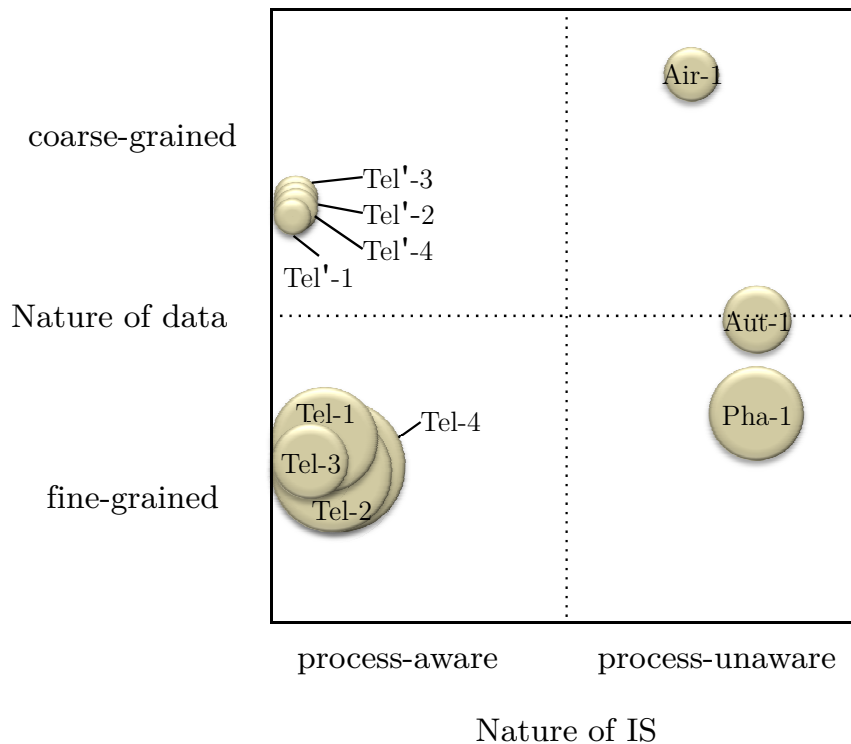


Figure 4.26.: Data preparation effort

p. 42), for instance, asserts that data preparation takes most of the time in any data mining project, estimating that this preparation can consume from 50 % to 70 % of the total project time. Pyle (1999, p. 16) goes further, believing it to be 60 % to 80 %. This benchmark shows that the effort required for process mining is kept within reasonable bounds, differing little with that of the classical data mining projects. That said, it could be argued that the very substantial high data preparation effort renders some data – notably enterprise data and RFID events – less suitable for process mining. However, from the viewpoint of the continual process improvement the following chapters will show that these process models provide valuable information and therefore are eminently suitable. And note, once the data preparation is established, it can be subsequently be executed repeatedly with much less loss of time.

The extent of process awareness and the level of detail of the underlying data are factors of the data preparation effort. Figure 4.26 depicts the important connections between these two with a four-quadrant model. The left quadrants are concerned with process-aware data, the lower one representing fine-grained actions, the upper one representing coarse-grained activities. The right quadrants are the domain of data, which is unaware of the underlying process, the lower one representing fine-grained events, the

4. Process Discovery with Process Mining

upper one representing coarse-grained transactions. The use cases Tel*, Tel'*, Air-1⁶⁰, Aut-1, and Pha-1 deviate considerably in their data and processes, as a result of which each of the four quadrants' dimensions were found in these use cases.

The data preparation effort saw increases from the lower to the upper quadrant and from the left to the right quadrant. As to the data's nature, the detail depth of a best practice activity serves as the point of reference, here represented by the dotted horizontal line. Take the observation of an RFID event. Since this observation is made without any knowledge of the process of which the event is part, this event belongs to the right quadrants. This event is in the right lower quadrant because the level of detail is below the point of reference. The circles' diameters are proportionate to the numbers of event classes – it follows that the circles' sizes indicate that the granularity of the data correlates with the different number of event classes – the higher the number the more fine-grained the data.

Attention is now turned to the process-aware use cases. Use cases Tel* and Tel'* are in the left quadrants. Since the WFMS supports the handling of cases in a structured way, these use cases show the greatest process awareness. Use cases Tel* employed the raw history of action, while use case Tel'* employed activities constructed from the history of action, a very important distinction. It took more effort to prepare Tel'* than Tel* because of this additional construction of activities. The data preparation effort increased only slightly because ProM offers powerful techniques for aggregating data, including filter techniques and data mining methods (e.g., clustering). The differences in data preparation effort within use cases Tel* and use cases Tel'* were minor and mainly due to data cleansing activities that were necessary because of varying data quality within the WFMS.

Now the less process-aware use cases are analyzed. In contrast to the use cases Tel* and Tel'*, appropriate tools for the preparation of the data of use cases Pha-1, Aut-1, and Air-1 were not available. Concept, design, and implementation of new ones result in high data preparation effort. The use cases Air-1, Aut-1, and Pha-1 resemble one another – each required a substantial effort as to both activity and case construction with varying degree of process-awareness. Use case Air-1 is more aware of processes than use cases Pha-1 and Aut-1 because compositions were involved in the process rather than aggregations.

The nature of use case Aut-1's data deviates from that of Air-1 and Pha-1. The granularity of events of use case Aut-1 corresponds to that of the activities of a reference model, and therefore required little effort to be prepared – indeed, much less effort than that required for use cases Air-1 and Pha-1. The transactional data of use case Air-1 is more coarse-grained. The RFID events of use case Pha-1 are more fine-grained. The decomposition of activities and construction of cases required high data preparation effort.

A four-quadrant model is also used to indicate that data preparation effort is not the only factor influencing the quality of the process models. For this purpose the fitness

⁶⁰ Since use cases Air-1 and Air-2 only differ in their data-time intervals, a difference that does not affect data preparation effort, use case Air-2 is omitted here.

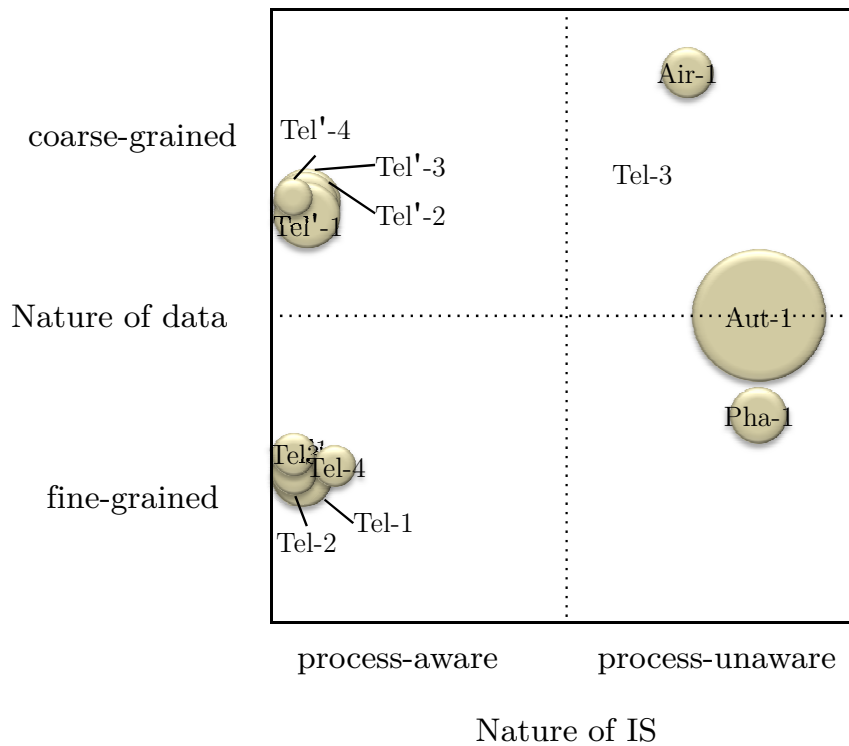


Figure 4.27.: Data preparation effort and process model's quality

of the resulting process models is proportionally used to create the diameters of each circle in Figure 4.27. The fitness values ranged from .28 to .99. Unfortunately, no benchmark reference values are available from the literature due to a paucity of industry examples. The percentages themselves patently indicate that the models in fact deviate very enormously. However, the circles' sizes show with the exception of use case Aut-1 that the models' fitness is between .28 and .49. In contrast, use case Aut-1 has a fitness of .99. The process model of this use case is a perfect representative of the simulated data because the simulation itself forbade variants of the activities, making it easy for the process mining algorithm to derive the dependencies among the activities. All evidence to date indicates that the process model's quality of operational data is also affected by a third dimension, that of process variance. The more freedom an employee has to execute activities the higher is the process variance whether or not the IS is slightly or fully process-aware.

Since in any given actual business situation there is virtually always more than one way to accomplish a process, and since it is not invariably clear which one is best, two or more employees tend to work differently, as a result of which process variance should not be considered noise but is quite understandably intrinsic to daily operation. In addition, the assessment of the process model's quality by each of the organizations in the case

4. Process Discovery with Process Mining

studies showed that in actual business situations intervals invariably deviated within a much shorter range, that is, between .28 and .49. This new, shorter interval demands a correction of the preliminary quality assessments – the models were of a moderate to relatively high quality rather than of a low-to-moderate one. Since the variance of the models' quality decreased from approximately .70 to .20, the quality of the models became more similar. These similar values of the operative use cases strongly suggest that the nature of both the data and the IS are relatively unimportant because data preparation practices can grapple with the problems inherent to data and to IS.

The second drawback must now be addressed. Note that process mining indeed projects the behavior of an IS onto that of an organization. Therefore, information outside the IS lies beyond the scope of process mining. Consider unrecorded activities such as manual activities and activities executed in an IS with no logging function – they are not recognized in the process model and hence may limit its expressiveness.

In sum, it is readily seen that process mining has its place, vis-a-vis discovering processes, along with the traditional method of manual process modeling. Both the objectivity and the automation of process mining are important features for maintaining transparency and for updating processes, things that are obviously valuable for measuring the congruity between how actual processes are performed and how they ought to be carried out, for identifying changes (such as variations in process execution), and for verifying that IT processes are in line with business requirements. Process mining has proven to be an important management tool for re-engineering efforts and it is not purely limited to process-aware processes. The time and effort needed to prepare data depends on the nature of the data and the underlying IS. There is a need for methods and tools that both integrate process mining as a tool in the continual improvement process and grapple with the difficulties identified.

5. Business Process Control Framework

This chapter⁶¹ presents an innovative approach leading to the genesis of the continual process and service improvement (CPSI) by the interconnection of the seven-step improvement process as part of the ITIL reference model and process mining. Based on the reference model, to-be processes are set up and key indicators are determined. As-is processes and their key indicators derived by process mining are compared to these to-be processes. This approach enables the design, control, and improvement of both business processes and ITSM processes – this will be demonstrated with real-world cases. This approach assays quality improvement vis-a-vis processes’ effectiveness *and* focuses on their efficiency. Ascertaining the most efficient use of limited resources in terms of personnel and tangible assets is an integral component.

Section 5.1 reviews the relevant literature. Section 5.2 explains the potential of the CPSI based on ITIL and process mining. Section 5.3 introduces a new framework for the CPSI. Section 5.4 turns from the more technical aspects of the CPSI to organization implementation. Section 5.5 concludes and adverts to further research avenues.

5.1. Related Work

This section incorporates other researchers’ views into the discussion of the process improvement. The extant sources can be grouped into three categories: analysis of event logs, process controlling based on measurement systems⁶², and improvement concepts in the data warehouse domain.

5.1.1. Event Log Analysis

Zur Muehlen and Rosemann (2000) devised the PISA tool that extracted performance metrics from workflow logs. The ARIS PPM allows for performance analyses of process instances (IDS Scheer AG, 2002) seeking deviations in KPIs to identify defective processes.

This paper relates to the mining of processes commonly employed in the business context. Case studies describing reverse engineering with process mining from various kinds of event logs in different application domains have been established recently. Mieke et al. (2008), for example, analyzed the procurement process in an SAP system for the purpose of internal fraud and theft risk reduction. Rozinat et al. (2009) investigated

⁶¹ This work is published in Gerke and Tamm (2009a) (*Prozessmanagement* in HMD, number 266, 2009), Gerke and Tamm (2009b), and Gerke et al. (2010a,b).

⁶² According to Fenton and Pfleeger (1996, p. 104), “measurement systems are used to assess an existing entity by numerically characterizing one or more of its attributes”.

5. Business Process Control Framework

feedback loops and idle times of a scanning test process to identify specific areas of improvement. They proposed iterative execution to enable continuous improvement. Mărușter and van Beest (2009) proposed a methodology comparing the mined model with a simulated model to predict potential performance gains with an eye for redesigning business processes. Segers (2007) integrated process mining in an audit approach of purchasing processes by testing control objectives with the linear temporal logic (LTL) checker.

Given the fact that few researchers have investigated the question of how to integrate the continuous improvement process with process mining techniques into BPM processes and ITSM processes, this integration has been little understood.

Furthermore this dissertation is related to process improvement and process compliance. Compliance has previously been discussed both in the context of business alignment (van der Aalst, 2005) and process redesign (Mărușter and van Beest, 2009). A promising approach for quality improvement in compliance is IT supported compliance evaluation (Sackmann and Kähler, 2008).

5.1.2. Process Controlling

Well-known controlling and performance measurement systems, such as the balanced scorecard (Kaplan and Norton, 1992) and Six Sigma (Brooks, 2006), support the evaluation and monitoring of processes to improve business processes. The process of building an objective indicator-based measurement system requires a thorough understanding of the relationship among processes, target values, and corporate goals – organizations face the challenge of determining relevant indicators. The approach herein provides guidance in selecting statistically significant key performance indicators.

5.1.3. Data Warehouse Concepts

Process improvement based on event logs is part of a broader thing, namely business process intelligence. Only a small number of authors, such as zur Muehlen (2001), Eder et al. (2002), and Casati et al. (2007), discussed the design of data warehouses taking advantage of event logs as an information source, and even works are limited to theoretical approaches or prototypical implementations. Due to challenges in storing and modeling the process warehouses, there are still issues requiring future research. The integration of business data is such a challenge.

5.2. Concept of Continual Process Improvement

Numerous organizations with intensive customer and user contact already use ITIL based processes for the optimization of their IT services⁶³. The issues and problems presented in the preceding sections can be dealt with through CPSI process by the interconnection of ITIL, the seven-step improvement process, and process mining. CPSI's potential is

⁶³ This is not only true for ITSM processes as will be shown in due course.

addressed and explained based on ITIL and process mining with the example of the incident management process, which is a central ITIL service operation process. Since the focus of this dissertation is on the continual improvement of processes, the life cycle phases *Service Operation* and *Continual Service Improvement* are described.

5.2.1. Procedure Model for Service Operation Based on ITIL

The user's interface is the service desk, the duties of which include receiving, processing, and monitoring of incidents and service requests. The goal of incident and problem management is to identify deviations from which incidents arise and to correct and to impose prophylactic measures. In this context incident management concentrates on restoring disrupted services as quickly as possible, whereas problem management focuses on substantial improvements to obviate future disruptions. Request fulfillment process is a separate process – it resembles incident management process but differs because it has separate request fulfillment records to handle service requests (Taylor et al., 2007a). Access management is the process of granting a prospective user the right to use a service. Event management monitors ostensibly normal operations to detect and escalate exception conditions. Service operation processes are accompanied by functions that sustain a stable infrastructure and a skilled workforce. Technical management provides detailed technical skills and resources to support the ongoing operation of the IT infrastructure. IT operations management executes daily operational activities. Application management is responsible for managing applications throughout their life cycle (Taylor et al., 2007a).

5.2.2. Procedure Model for Continual Service Improvement Based on ITIL

The volume *Continual Service Improvement* calls for continual improvement of IT services. In order to control both IT services and IT processes, it is crucially important to know precisely *what* needs to be measured and *why*. Verification of key goal indicators (KGIs) and target values needs to be performed initially to determine whether process goals and associated effectiveness will be reached (IT Governance Institute, 2007). Key performance indicators define measurement ranges, which, with trend data and benchmarks, indicate whether or not process performances are adequate, for if they may not be, reaching a process goal and its associated efficiency is jeopardized. Key indicators include the number of incidents and the reaction time. The perpetual juxtaposition between the to-be condition and the as-is condition is delineated in seven steps (Taylor et al., 2007b):

1. Define what should be measured: Identify criteria and goals through the design of processes, while simultaneously verifying the process with respect to quality, performance, and compliance.

5. Business Process Control Framework

2. Define what can be measured: In the framework of given obligations, the relevant boundaries, that is, IT resources and available budgets, are identified from the requirements of business processes.
3. Gather the data: Data aiding in the identification and proving of causes for deviation is collected.
4. Process the data: Data has to be converted to a consistent format to compare findings from a variety of sources.
5. Analyze the data: Key indicators are integrated as measurement points in the process management and analyzed periodically while always displaying them as a trend and in contrast to the to-be values. In the framework of such benchmarking, a common basis for comparison needs to be ensured.
6. Present and use the information: Necessary corrective actions need to be communicated to the organization and subsequently analyzed according to cost-benefit factors and probable results.
7. Implement corrective actions: A comprehensive implementation plan is devised and implemented based on the ITIL volume *Service Transition*'s recommendations. Subsequently, the improvement process starts anew.

5.2.3. Procedure Model for Service Operation and Continual Service Improvement Based on Process Mining

Figure 5.1 shows the procedure model based on process mining. Many organizations use ISs, which are configured based on to-be process model M_2 (see 2), for the support of their ITSM processes. Process models formally describe business processes.

Reference model M_1 provides preliminary clues for the implementation of these processes (see 1). At the beginning of a process execution an instance is decided that may consist of various activities (see 3), such as the receipt of a user's incident. Generally, the execution is recorded through the IS and saved in event logs (see 4). The figure shows that all instances are recorded with the exception of process P_b . Log file L_a only contains activities A and B since activity C is manually executed. For the formalization of the event logs, MXML, which is required by ProM, is used. Based on event log L_u , the process mining engine derives the intrinsic present knowledge (see 5) in an as-is process model M_3 while considering key indicators and goals. The evaluation engine (see 6) compares the as-is processes M_3 with both model M_1 and model M_2 , to continually determine the degree of M_3 's compliance. Therefore as-is processes are probed for weaknesses and potential sites of improvement. Furthermore, being integrated into the procedure model, maturity models determine the quality of the customer support process and provide action recommendations to improve process quality. According to each level of abstraction, an adaptation on one or more of the following levels may be necessary (Gerke and Tamm, 2009a): model M_1 , model M_2 , and instance.

5.3. Definition of a Business Process Control Framework

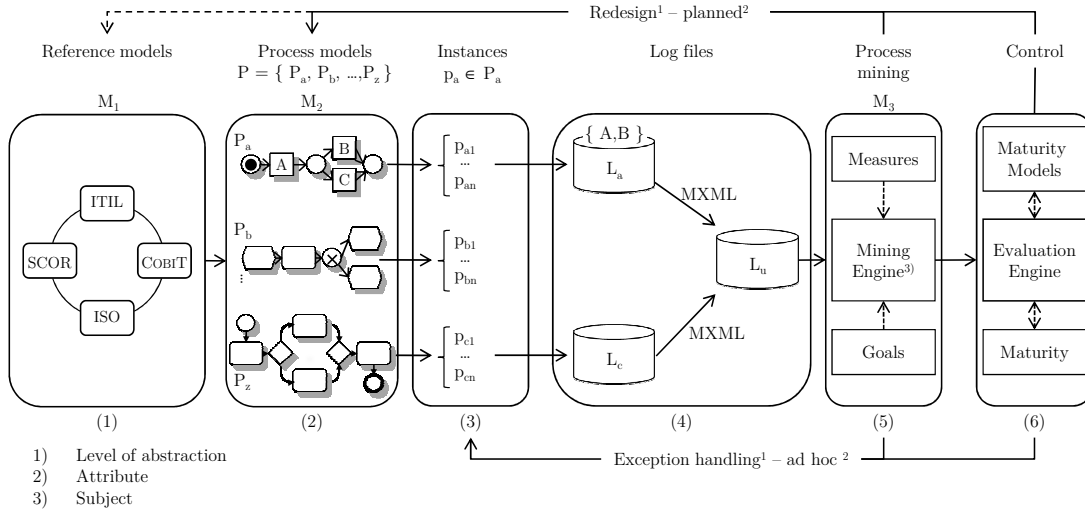


Figure 5.1.: Procedure model for service operation through process mining (Gerke and Tamm, 2009a)

5.3. Definition of a Business Process Control Framework

This section establishes the need for a new framework. Section 5.3.2 describes comparable entities. Section 5.3.3 introduces a procedure model for IT Service Operation based on ITIL and process mining. Section 5.3.4 shows that the procedure model is compatible with the reference model COBIT.

5.3.1. Need for a New Control Framework

This section establishes the need for a new framework from two different points of view: internally motivated need and externally imposed need. The former is established by analyzing the present situation of Internet service providers⁶⁴ – the latter by elucidating externally imposed business criteria on process management. This need will be reinforced by insights gained through a series of expert interviews.

In today's highly competitive environment, IT service providers are judged on their ability to deliver contractually specified services in a timely manner – pricing competition is in no small measure part of this, thereby putting pressure on providers to reduce total costs of ownership in order to be able to offer their services at an acceptable cost-benefit ratio. Equally important, the ever shifting business environment ineluctably increases the frequency, complexity, and extent of change, thus compelling customers' demands for the utmost flexibility as to IT services and for their requirements as to ITSM.

⁶⁴ This is not only true for ISPs as will be shown in due course.

5. Business Process Control Framework

Because of these pressures, cost reduction and quality enhancement have become tremendous factors. It is important that this commitment to quality and cost is targeted not only to production processes of IT services, but to IT service management activities.

To deliver and maintain services in a prompt, cost-effective, and reliable manner, organizations have been advancing the automation of their ITSM processes and making use of emerging technologies. Note that for years a great deal of effort has been put into improving process quality – it is as important to develop further the continual improvement process itself. Both an efficient process management’s continuity and cost effectiveness requires this. And the automation of both the operational processes themselves and process management activities is essential for high maturity levels (The IT Governance Institute, 2007). Process mining, for example, facilitates the analysis of processes by extracting a process model from event logs and has considerable potential of automation. In addition to which, IT service providers make use of reference models. ITIL, for instance, is among the most commonly used frameworks of service delivery (IT Governance Institute, 2008a); it provides guidance on creating and maintaining value for customers through better design and operation of services. ITIL can be used with current methods and tools, especially since ITIL cognoscenti invariably recommend that organizations adopt ITIL processes within their own context. Therefore the question arises how procedures and emerging techniques, such as process mining, can be seamlessly integrated into ITSM processes to improve continually ITSM processes and services.

To optimize the delivery of IT services to customers and users, two additional things must be addressed. The first issue is the most efficient use of limited resources in terms of personnel, systems, and resources. The second issue arises from the fact that different services share the same IT process. A little consideration will show that it needs to be proved if these processes hold for service-specific peculiarities.

After the processes have been designed or redesigned according to the reference model, it is necessary to check continually the process execution. In order to identify quality problems, organizations commonly measure the efficiency and effectiveness of their ITSM processes with key indicators. Target value verification allows analysis to determine if a process goal’s attainment is in jeopardy.

Setting up such a measurement system involves definition of indicators relevant for the purpose in question – these definitions are difficult for two reasons. First, no organization knows the optimal set of indicators in advance – and the attempt to specify before-the-fact results in a selective monitoring process that inevitably limits control and improvement opportunities, leaving key relationships unmonitored and even hidden. Second, the metrics for process monitoring should adapt as the strategy or goals change. Morgan and Schiemann (1999) have stressed that metrics that are outdated or those lacking alignment with organizational objectives may even obstruct the benefits.

Since process mining aims to reveal process information, one is compelled to ask, if this capability can be used to propose performance indicators that themselves may lead to insight for process improvement.

To grapple with the above challenges, ITSM cries out for a control framework that facilitates the continual process improvement.

5.3. Definition of a Business Process Control Framework

In addition to these internal motivated needs, the IT service provider must consciously deal with externally imposed business criteria to ensure that its business processes are executed as they ought to be and to ensure that both operations and practices are consonant with all laws, regulations, standards, and practices. Consider risk management. In the literature, collaboration between process management and risk management is vehemently urged.

Menzies (2004, pp. 183 ff) believes that the analysis of process models is an appropriate means to identify risks that are subject to the KonTraG and the SOX. The Committee of Sponsoring Organizations of the Treadway Commission (COSO) (2004, p. 26) has written to similar effect and employs process model analysis for the risk management in its reference model COSO. Basel II includes processes in the risk definition (Basel Committee on Banking Supervision (BCBS), 2006, p. 137). Dörner (1998, p. 2), with reference to KonTraG's introduction, explicitly opines that the business processes and the business environment of the organization are the starting point for risk appraisal, *not* the annual report and accounts. In order to ensure compliance with these mandates, an IT service provider needs a process control framework to map abstract compliance requirements to concrete control structures and processes, to enforce controls in business operations, and to evaluate effectiveness of controls.

The above cogently militated in favor of the need for a process control framework. This was verified empirically – experience surveys were conducted with eleven chief information officers (CIOs) with relevant ITSM and BPM expertise from industries not represented by the use cases' organizations. All surveys were conducted in Germany between September and November 2010. Although extrapolation from surveys and interviews suffers from the intrinsic unreliability of generalizing from small samples, research surveys are a widespread instrument of qualitative research (Wilde and Hess, 2007). It has been chosen because of the research's exploratory nature. A number of questions vis-a-vis the importance of IT management deliberately resemble those asked by the IT Governance Institute (2009), which conducted 255 interviews to analyze the view of practitioners regarding IT governance. The congruence between the results strongly indicates that those queried for this dissertation were not atypical. Like ITGI's survey, the survey herein indicated that IT was important. Nearly half of the respondents of both surveys felt that IT was very important to the organization. The surveys analyzed the view of practitioners as to IT management activities in light of IT's contribution to business process improvement. A discussion regarding this need follows⁶⁵.

IT's contribution to efficiency, innovation, and strategy was widely accepted (see Figure 5.2). More than half of those surveyed felt that IT was very important to the organization. IT's contribution to business efficiency was deemed more important than its innovative value. More than half of the number of respondents considered IT very important to the organization's ability to achieve its strategy, with one-third noting a very important contribution to IT vision. Almost three-quarters of those surveyed indicated that their IT management maturity was between levels two and three, meaning that processes had developed that were either repeatable due to similar procedures, though

⁶⁵ See Appendix A for the questionnaire and its analysis.

5. Business Process Control Framework

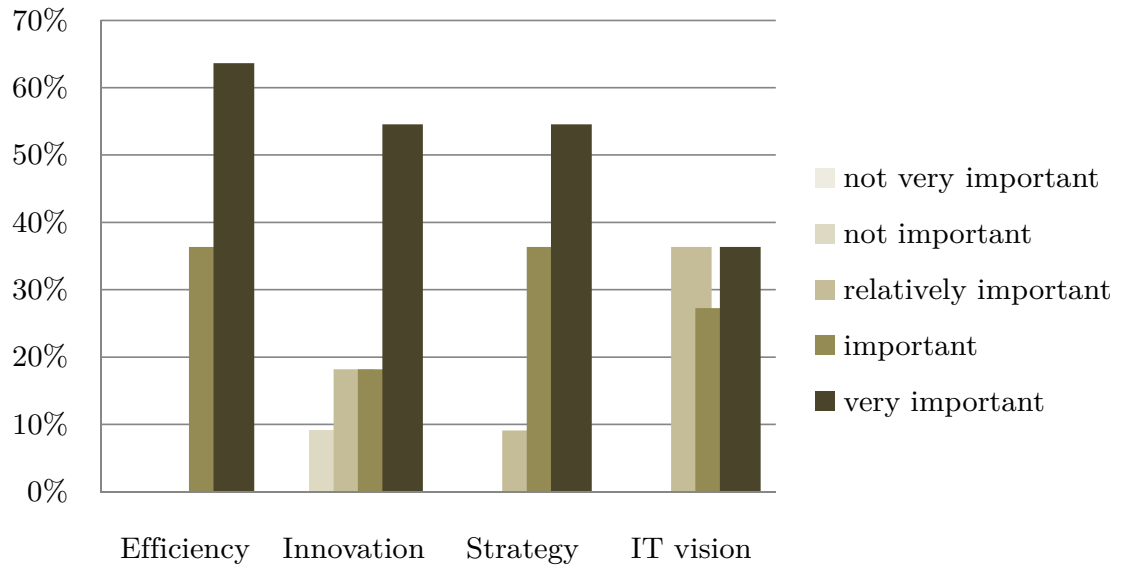


Figure 5.2.: IT's contribution to efficiency, innovation, and strategy

intuitively managed – level 2 – or that were standardized, documented, and communicated – level 3 (see Figure 5.3). Nearly a fifth indicated that their management processes were managed and measurable (level 4).

All respondents indicated that there were challenges complicating organizations' management efforts (see Figure 5.4). More than four-fifths experienced very substantial pressure from business needs (e.g., time-constraints, frequency of changes, and cost), nearly two-thirds felt themselves to be confronted with the high speed of IT innovation, poor business requirements, as well as organizational aspects, and nearly half found the absence of appropriate IT management methods to be waxing.

When asked about the effects of IT management, the respondents were unanimously positive (see Figure 5.5). Increased process transparency, clear ownership and respon-

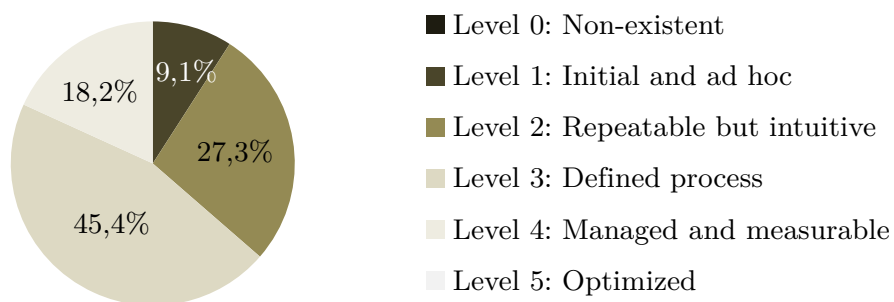


Figure 5.3.: Maturity of IT management

5.3. Definition of a Business Process Control Framework

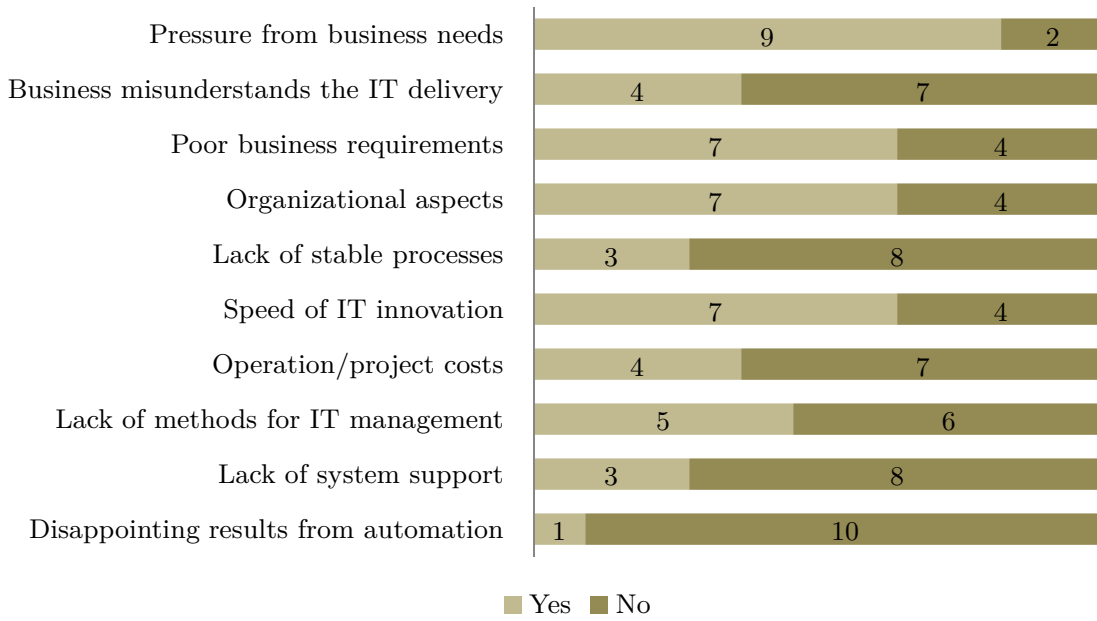


Figure 5.4.: Challenges in IT management

sibility based on process orientation, increased IT efficiency, and enhanced IT-Business alignment were the salient benefits identified.

As-is processes were analyzed exclusively by traditional methods, that is, interviews

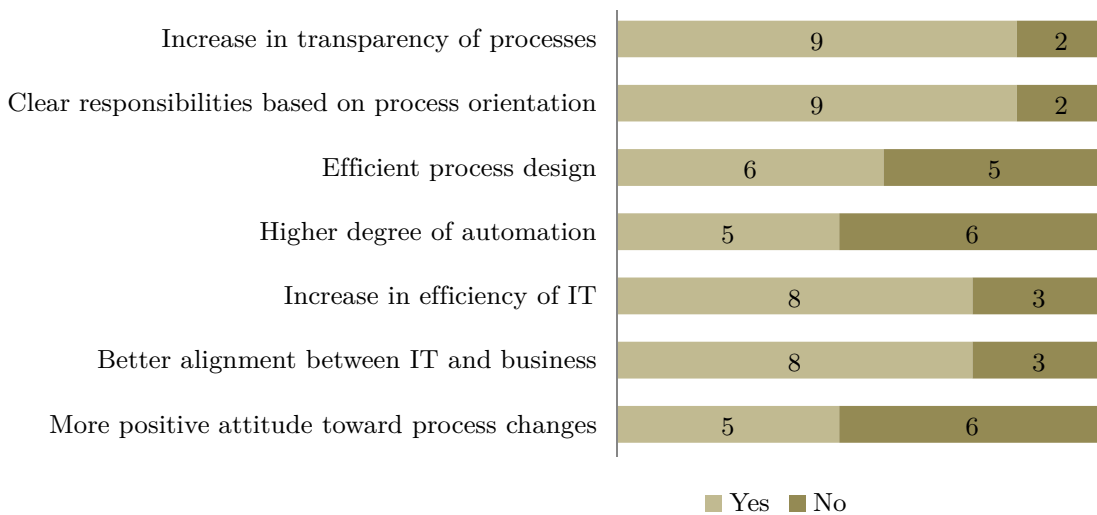


Figure 5.5.: Outcomes of IT management

5. Business Process Control Framework

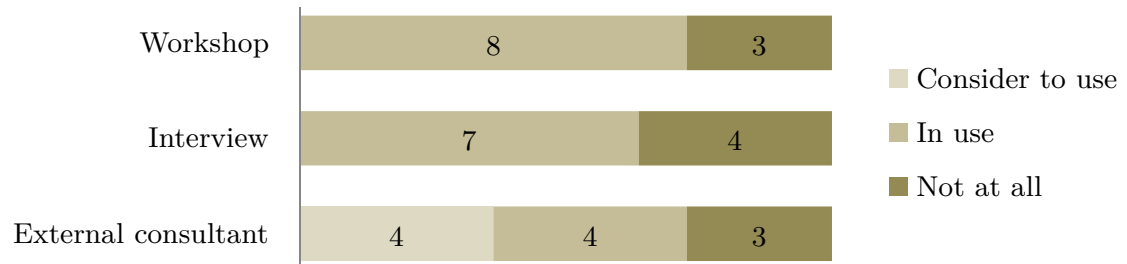


Figure 5.6.: As-is process modeling

and workshops. More than one-third of the respondents said that they called on external consultants for process modeling and another third considered their use (see Figure 5.6).

With respect to process performance (see Figure 5.7), more than half of those surveyed failed to analyze this performance or did so sporadically and infrequently, doing it no more than one time each fourth month. However, given the fact that more than 60 % of respondents put the processes' maturity level between three and four, there is a blatant inconsistency – a number of insouciant respondents ought to have been concerned with process compliance specifications (and with continuous process improvement), but clearly were not. The presence of a more sophisticated performance measurement, that is, performance measured at least every two months, was indicated only by 11 % surveyed vis-a-vis process compliance and by 20 % of those vis-a-vis process improvement. These surprisingly low values strongly hint at a potential gap between the maturity perceived by the organizations and that defined by the reference model.

Though over 82 % of those surveyed indicated that they were personally aware of

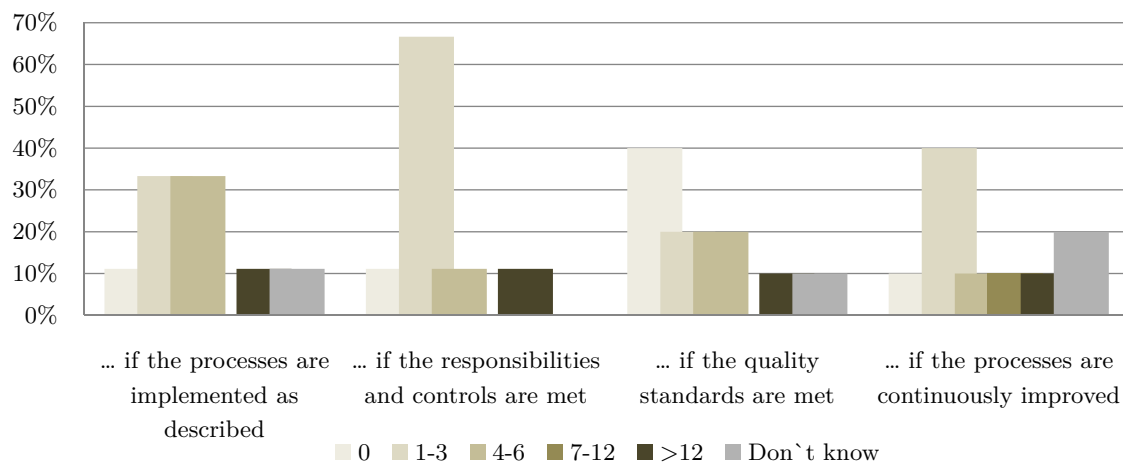


Figure 5.7.: Process performance analysis

5.3. Definition of a Business Process Control Framework

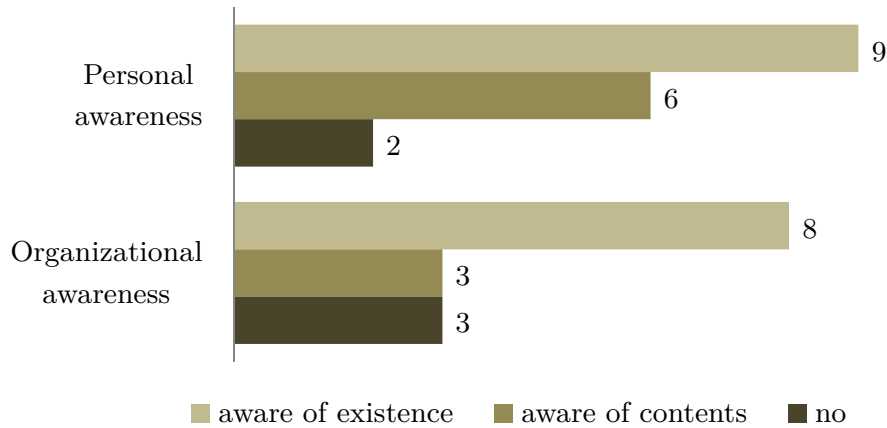


Figure 5.8.: Awareness of reference models

reference models, with 55 % aware of the content (see Figure 5.8's upper part), less than 50 % instilled their knowledge into the organization so that only 27 % knew the content (see Figure 5.8's lower part). Among the respondents, ITIL and ISO 2000 were used most frequently, by over 43 % – virtually all of the others indicated they would consider their use. Well-known reference models included the ISO models, ITIL, Capability Maturity Model Integration (CMMI), COBIT, Val IT, and COSO (see Figure B.9).

Expectations regarding the use of reference models varied. There was roughly an equal level of enthusiasm⁶⁶ (about 90 %) about anticipated benefits from reference models vis-a-vis IT-Business alignment, business process improvement, IT process improvement, and time or cost reduction (or both of them) for process modeling (see Figure 5.9).

Regarding organizational aspects and IT compliance, 70 % of the respondents expected benefits.

Although only 36 % (see Figure B.14) of the respondents had been aware of process mining prior to being surveyed, they after being appraised, overwhelmingly indicated that process mining would be helpful for their companies vis-a-vis process analysis, process improvement, and compliance with reference models.

Business processes were expected to benefit from compliance with reference models – both IT processes and business processes were expected to benefit from process analysis and process improvement (see Figure 5.10).

The surveys have indicated that the overwhelming majority of the organizations proactively managed their IT functions and services, that is, they oversaw them with considerable maturity – indeed, this management was perceived as very important – though the overall maturity was, strictly speaking, moderate. These organizations have affirmed that IT management has created business value, with the proviso that major challenges, such as pressure from business needs, speed of IT innovation, and absence of IT management methods, persist in making IT management difficult. Surprisingly, more than half of those surveyed has failed to recognize that process performance, especially compli-

⁶⁶ Note that those who agreed or strongly agreed are considered.

5. Business Process Control Framework

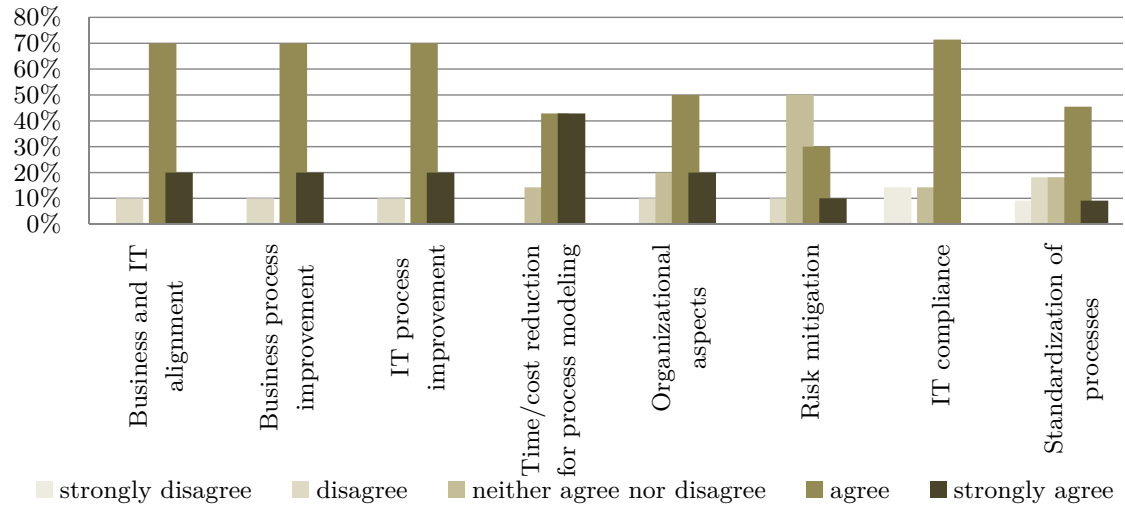


Figure 5.9.: Expected benefits from reference models

ance with specifications and continuous process improvement, is an issue that needs to be addressed. Despite the fact that reference models' utility had been broadly accepted, their penetration in business was limited due to barriers presumably linked with challenges inherent in IT management. Process mining, even when not actually employed in practice, was seen as helpful in advancing process improvement. The importance of IT generally, those challenges to its management, and impediments to a wider use of both reference models and process mining all need to be addressed.

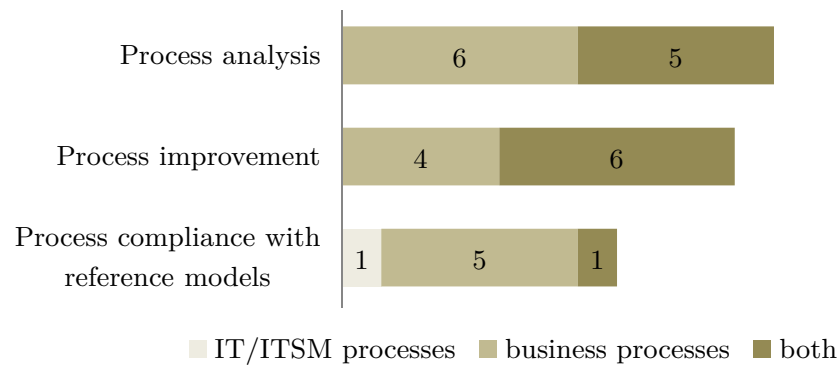


Figure 5.10.: Usefulness of process mining

5.3.2. Relevant Entities

While traditional approaches to gauge compliance only consider measures referring to the execution of business processes (e.g., first kill rate), the adoption of reference models

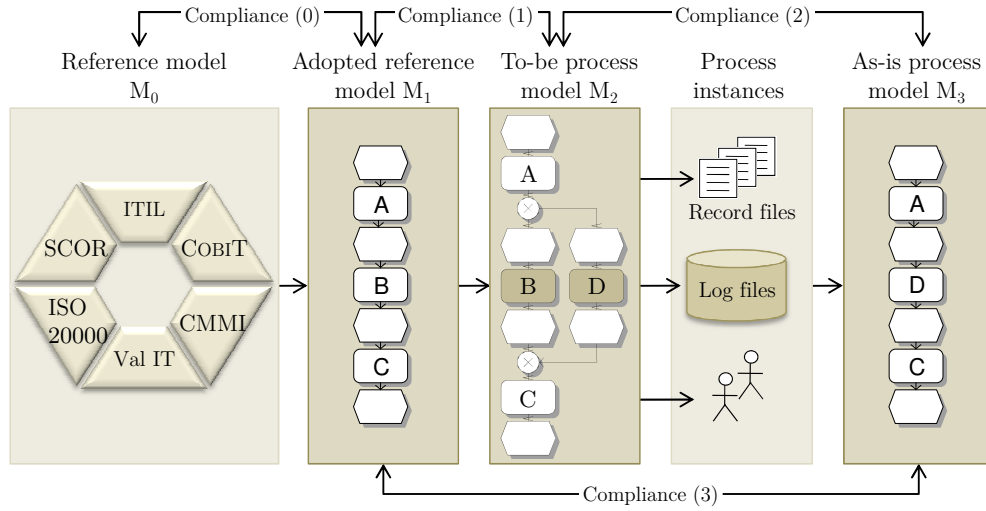


Figure 5.11.: Entities of a compliance analysis

brings new entities into the modeling task. The methodology identifies five entities, illustrated in Figure 5.11 that need to be considered when measuring the compliance with reference models: the meta reference model M_0 , the adopted reference model M_1 , the to-be process model M_2 , model M_2 's instances, and the as-is process model M_3 . Depending on the scope, model M_0 provides either generally accepted processes or a set of abstract guidelines. In both cases, and particularly in the latter case, the reference model M_1 needs to be adapted to the needs of an organization to yield a set of processes M_2 . This set incorporates the vision, goals, and objectives of the organization and can consider future processes. The processes' execution generates a set of instances. The analysis of these instances provides model M_3 reflecting how a process of model M_2 was executed.

The level of compliance can be measured by analyzing process models M_0 , M_1 , M_2 , and M_3 . Given the fact that model M_0 is typically specified in the vernacular (in laymen's words), the CPSI concentrates on analyzing models M_1 , M_2 , and M_3 . Models M_1 and M_2 are usually constructed manually, whereas M_3 is usually inferred from event logs, that is, recorded process instances.

5.3.3. Procedure Model for IT Service Operation Based on ITIL and Process Mining

Figure 5.12 represents a two-phase approach to improve continually ITSM and business processes. The use of process mining and the application of the ITIL-recommended seven-step procedure are essential to the approach. It pursues two major aims – it identifies and corrects deviations from reference processes and it determines and rectifies service-specific weaknesses of the process implementation. The approach integrates pro-

5. Business Process Control Framework

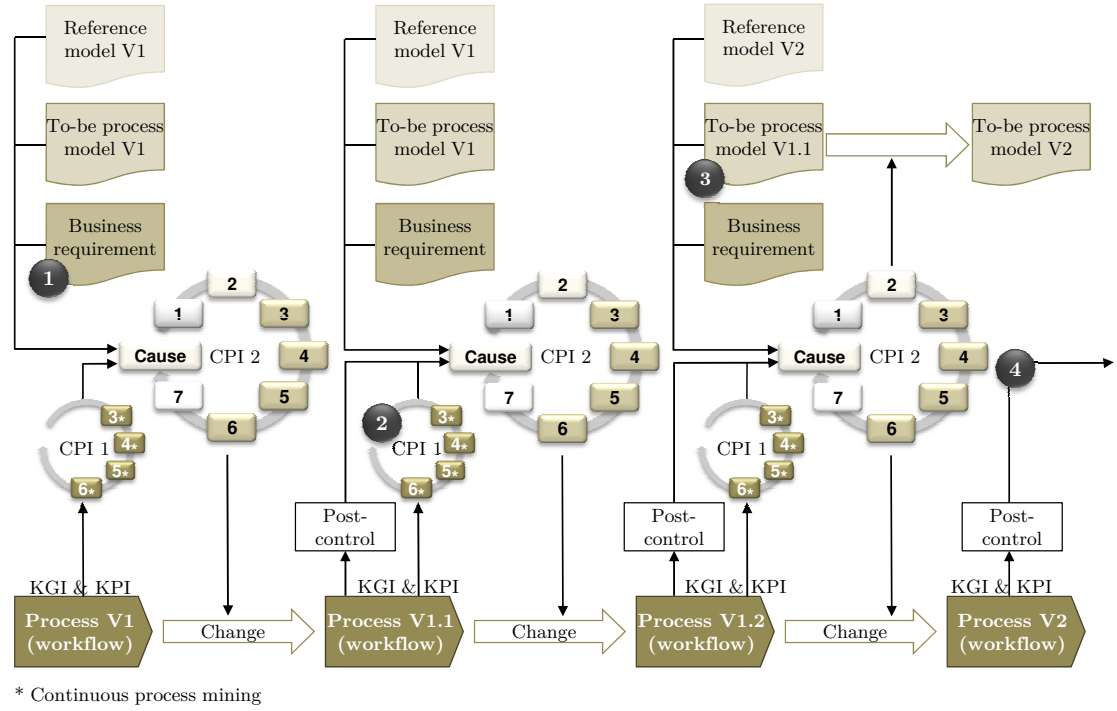


Figure 5.12.: Continual process improvement approach

cesses, personnel, systems, and resources. The following describes how compliance with reference models is ensured.

In the first phase, each ITSM process is continuously monitored as part of service operation processes. Operational monitoring and control ensure that the ITSM process functions as specified. The first control cycle (CPI 1) therefore is keenly concerned both with the verification of key figures (key performance indicators and key goal indicators) against target values and with the compliance of as-is processes with to-be processes. This control cycle inherits steps three to six of the seven-step procedure. All steps are supported by process mining – all steps automatically measure and compare, and they alert one to the meeting of to-be specifications. It follows that this control cycle contributes to a continual inter-departmental and inter-organizational process approach to quality improvement of ITSM processes. Once the process identifies a possible deviation the second phase is triggered – this, the CPI 2 is a continual⁶⁷ improvement approach that can be applied in a semi-automated way. CPI 2 passes through all steps of the underlying seven-step procedure. Four major changes can initiate the second phase of the CPSI.

⁶⁷ Continual and continuous is indeed different. Continual typically marks a close and unbroken succession of activities, rather than absolute continuity. Continuous is the stronger word, denoting that the continuity or union of activities is absolute and uninterrupted (Taylor et al., 2007b).

5.3. Definition of a Business Process Control Framework

1. Changing business requirements might entail adapting the design as well as the underlying implementation of model M_2 – it could be that a service is too costly or it could be that the business needs to change the service portfolio or both of these. Changing may redefine key indicators and their target values. Consider rising expectations as to the first kill rate – they might result in the need for additional tool support and, obviously, in higher target values for the KPI.
2. These changes can also be initiated by the identification of deviations both between key indicator values and their target values and between model M_3 and model M_2 . Returning to the above example, the continuous control of the adapted target values may reveal the chronic under-performance of the first kill rate.
3. A new version of the reference model can also trigger the above changes. The segregation of incident management and event management in ITIL v3 of the incident management process provides one recent example.
4. The approach supports the post-control of measures taken in a business context. For example, the approach can detect whether or not a workflow step thought to be removed by an order is nonetheless extant. Deviations can appear because the corrective actions are not appropriate or executable, or simply because of inadequately trained staff. Post-control enables the process manager to revise or augment a given measure.

To contemplate the optimization of the delivery of IT services to customers and users, two additional things must be borne in mind. First, one must consider the most efficient use of limited personnel, systems, and resources. Second, one notes the troubling fact that different services share the same IT process – the question arises whether such a process is proper for service-specific peculiarities.

Examination of the use cases (about which more in due course) has revealed that it is *not* true that deviations are not alike. A number of deviations can be explained by peculiarities inherent in the services, whereas others stem from deviant working behavior. Evidently, vis-a-vis deviation patterns, there is a distinction between those stemming from weak points that are caused by the process itself and those stemming from weak points as a result of the process implementation. The former is referred to as non-adherence to references (NAR) and the latter is referred to as adherence to references (AR). As the CPSI approach is currently only aware of structural deviations, namely pattern NAR, it has been extended as depicted in Figure 5.13.

Step five of the control cycle CPI 1 includes a pattern analysis that determines whether a problem lies within the process or within the process execution. As to pattern NAR, the process deviates from the specification, as a result of which the CPSI approach proceeds as it was initially meant to be.

The determination of pattern AR triggers the control cycle CPSI 1. The rationale behind this cycle is that the process does optimally support the business but has to be improved service-specifically with respect to actual implementation. The semi-automated analysis has been done by control cycle CPSI 1, consisting of all steps of the seven-step

5. Business Process Control Framework

procedure. Note that the analysis of the deviation (i.e., steps three to five) has been done in an automated way. It is supported by means of process mining and results in a measurement system that in tandem with the process model supervises the complete process course and identifies statistically relevant KPIs. Therefore they need not be determined in advance.

The KPIs can be analyzed with respect to distinctive features of the process to detect the sources of the deviations. Examples of possible differentiators are services, spheres of responsibility, cooperation models, and resources. If a deviation is due to process execution, two potential solutions present themselves. First, it may not be necessary to change the process itself – introducing a new resource, such as engaging additional staff, or improving a service-specific activity, such as enhanced staff training, might be efficacious. The activity changes the resource set in terms of personnel, systems, and resources. Cycle CPSI 1 continually repeats itself until the performance indicator is within the normal range again, that is, until an efficient resource set with respect to the internal benchmark is found. Second, one may simply accept the putative deviation because it is specific to the differentiator, in which case the comparison base (i.e., to-be model) for process mining has to be adjusted or augmented.

The two-phase approach shows that the quality of the ITSM processes substantially depends on how well they are monitored and measured in service operation. This approach can be consummated through a variety of concept of measurements including Deming cycle (Deming, 1982) and Six Sigma (Brooks, 2006).

5.3.4. Compatibility with Reference Models

Since the development of the CPSI rests on ITIL, the compatibility of another reference model, COBIT, is now demonstrated.

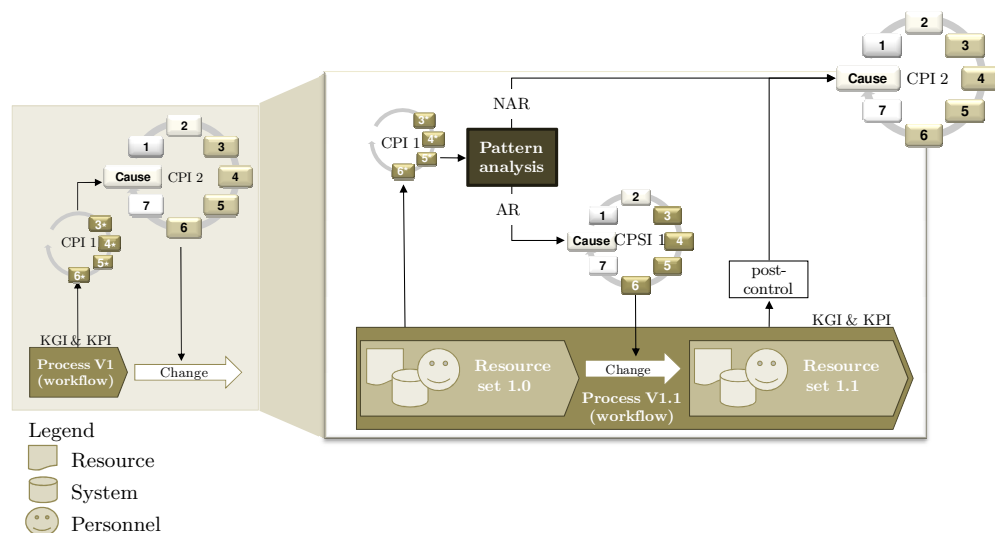


Figure 5.13.: Extended CPSI approach

5.3. Definition of a Business Process Control Framework

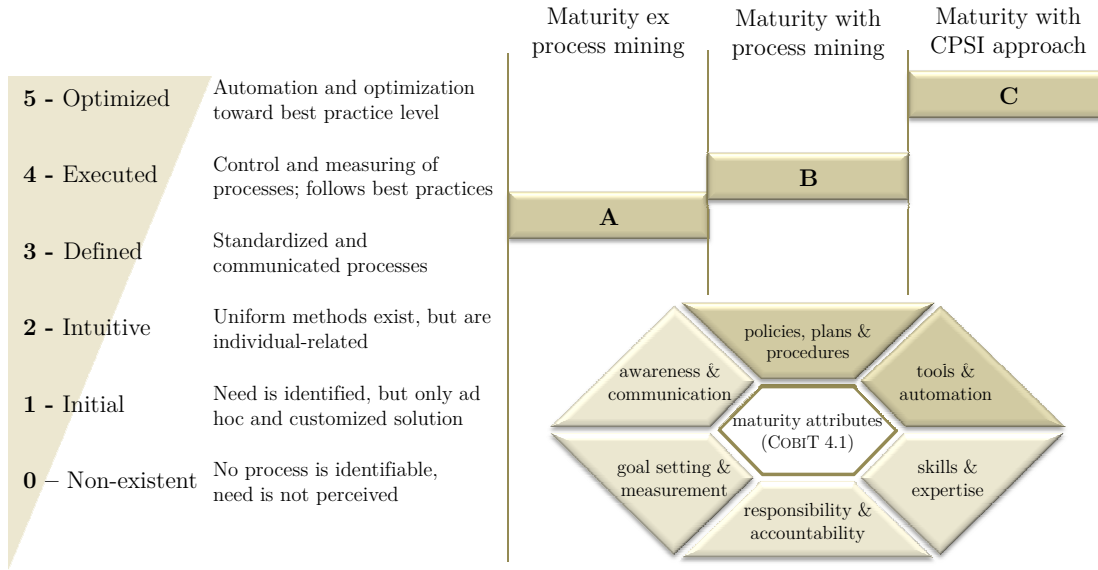


Figure 5.14.: Maturity assessment of the incident management

As noted above, reference models are often used with other best practices compositely. One therefore must determine if the integrated use of process mining within CPSI is suitable for improving the maturity level of the ITSM processes from the view point of COBIT. Figure 5.14 depicts the assessment of the maturity of the ITSM process *Incident Management* both with and without process mining, and finally with the CPSI approach.

The maturity model consists of six maturity levels (0-5). The placement into a level is based on six attributes of maturity (IT Governance Institute, 2007), which are shown in the figure's hexagon. Consider, as examples of the influence of the attributes on the classification, the two attributes *policies, plans & procedures* and *tools & automation*. According to the maturity attribute *policies, plans & procedures*, maturity level four is reached if model M_3 complies with internal best practices, in which case the internal best practice is represented by model M_2 (case A). If both the internal *and* the external best practices are applied, the maturity level raises to five (case B). The external best practice might be represented by the ITIL reference model M_1 . Thus, the combination of process mining with ITIL is crucial for compliance with maturity level five. The maturity attribute *tools & automation* requires tool use for maturity level four for process management and monitoring that is addressed by applying process mining. The requirements of maturity level five are indeed rigorous, demanding as they do process improvement and automated detection of control exceptions – yet the application of the CPSI approach fulfills even them (case C).

5.4. Continual Process Improvement in Practice

Thus far, this chapter has been examining the CPSI from a conceptual perspective. This section both takes a close look at the CPSI in corporate practice and describes its application as to use cases⁶⁸ Air-1 and Tel* that had suffered from deficiencies in their then-extant processes, and illustrates the benefits these organizations obtained from the CPSI based on ITIL and process mining. The industrial application evidently proves the concept.

5.4.1. Complaint Handling

This section outlines the practical relevance of the CPSI approach by means of use case Air-1. The focus of the improvement project was on the compliance of the as-is complaint handling process with specifications.

1. Define what should be measured. A successful management of customer relations was part of the airline's strategy. The enumerable indications contained in customer complaints pointed toward quality deficiencies the potential rectification of which could result in increased customer satisfaction. This aim was closely tied to the goal of optimizing the complaint processing to simplify and standardize complaints. The following had to be measured:

1. Compliance of business process *Complaint Management* with ITIL.
2. Maturity of IT processes.

2. Define what can be measured. Three processes were available: an ITIL based reference model M_1 ; a to-be model M_2 including future processes, such as a planned online complaint collating system; an as-is model M_3 mirroring the current complaint processing in the CRM system.

1. Statements about compliance were derived from the comparison between two models respectively (henceforth " (M_x, M_y) "). Comparison (M_1, M_2) provided insight pertaining to the degree of ITIL compliance. Comparison (M_1, M_3) led to further analysis of the technical maturity of the IT processes. Comparison (M_2, M_3) evaluated the service desk.
2. The maturity degree was evaluated with the aid of the maturity model based on the COBIT process *Manage Service Desk and Incidents* (see Figure 5.14). From the results of the compliance analysis, the as-is condition can be derived and subsequently contrasted to the to-be condition. Action alternatives for the improvement of the business and IT processes can be deduced from discrepancies.

⁶⁸ Use case Pha-1 information is unavailable. Since use case Aut-1 suffered from no deficiencies, it is omitted here.

3. **Gather the data.** The relevant processes for model M_1 (see left-hand side of Figure 5.15) were manually modeled using an eEPC according to the conceptual guidelines of ITIL. Various activities include *create incident*, *categorize incident*, and *prioritize incident*. Model M_2 was similarly conceived in coordination with process experts of the airline (see the middle of Figure 5.15). The incidents were handled with an CRM system that stored relevant information in its database.
4. **Process the data.** Process discovery of the complaint handling was employed as described in Section 4.2.3. The resulting model M_3 is depicted on the right-hand side in Figure 5.15. Prior to the compliance analysis, the process models were converted to Petri nets through a conversion plug-in available in ProM. The mapping among the events was manually done because events were present in different granularities.
5. **Analyze the data.** The ProM plug-in *Conformance Checker* allowed the quantification of the compliance of one model with a second. Figure 5.16 depicts comparison (M_2, M_3). Positive values in the circles show how often activities that

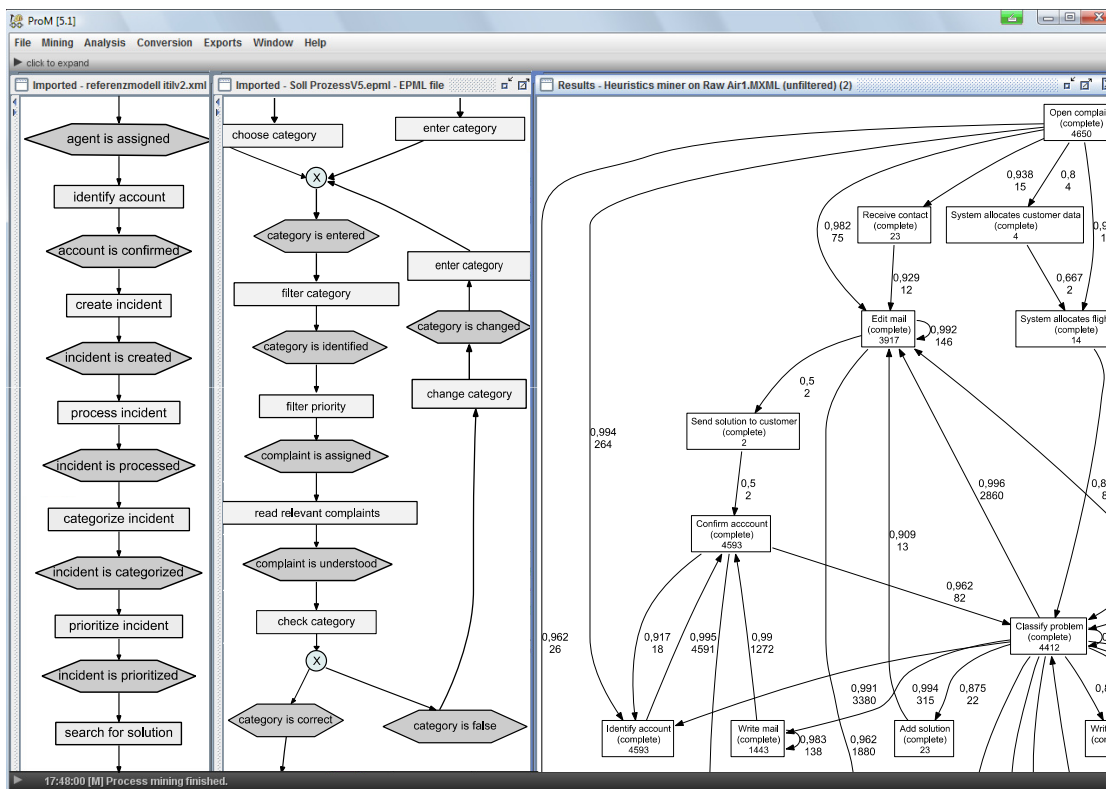


Figure 5.15.: Model hierarchy Air-1 for complaint management

5. Business Process Control Framework

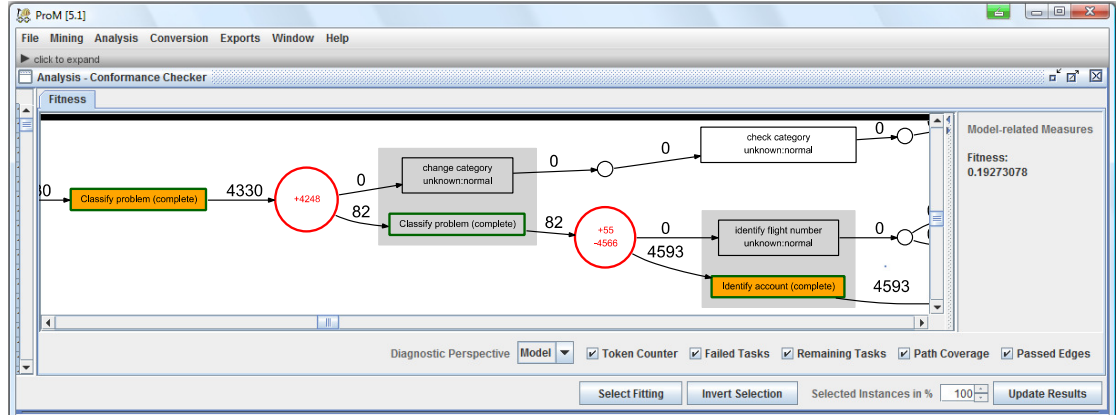


Figure 5.16.: *Conformance Checker* plug-in of model Air-1

should have been executed were (heavily-shaded); negative values show how often activities have been executed though unplanned (lightly-shaded).

Activities displayed in white are absent in the event log. The number of edges denotes the complete number of executions. Take activity *Identify account*. It was executed 4,593 times, in 55 of which a classification was not performed. In 4,566 cases, a classification was performed although this activity was not planned (-4,566). Subtracting executed activities from skipped activities and adding those that were performed in deviation to the model, the sum of executed predecessor activity is arrived at: $4,593 - 4,566 + 55 = 82$. The fitness of comparison (M_2 , M_3) was .19, that of comparison (M_1 , M_3) .39, and that of comparison (M_1 , M_2) .25. Inverting the sequence of two activities in the reference model showed that the fitness changed substantially, from .39 to .60. Note that, herein, fitness values will serve as reference values for further comparisons. The compliance analysis is likely to yield higher fitness values as the technical maturity of the IS increases. The comparison provided valuable information. Since model M_2 incorporated activities that were planned for the future, a starkly large number of white activities were present. It follows that a large number of these activities were at that time performed manually – fortunately, they will be instead performed by the planned online processing system. The analysis also showed that complaints were not neither routinely prioritized nor forwarded, and both whether and how a solution was being provided was not routinely reported. These results were incorporated into the maturity model as initial estimates of the as-is situation.

- 6. Present and use the information.** The necessary measures to increase the quality level were determined within an airline workshop. This new-found transparency over the process execution had positive consequences. It made standardization of the complaint processing feasible. And the integration of process mining allowed for better control of IT services and helped to ensure business processes' compli-

ance. The airline found the low values of compliance seemingly inexplicable. An explanation follows. The above experiment clearly showed that the ordering of activities is a crucial aspect influencing the outcome of the compliance analysis.

- 7. Implement corrective actions.** An adaptation of the application IC was done to record exact time stamps for all activities associated with a complaint. The CRM system was adapted to compel IC agents to enter more details of customer interaction.

5.4.2. Service Operation

This section analyzes the expediency of the CPSI approach with the use cases Tel* to its full extent. The data available allowed both an analysis determining whether or not ITSM process *Incident Management* was executed as specified and the comprehension of the effects that different process variants (services) had on ITSM processes and the influence of personnel, systems, and resources.

- 1. Define what should be measured.** The management of incidents posed a serious challenge because the ISP had to be restrained from having a negative impact on user experience. Once a customer or a user voiced an incident, the service operation processes had, by definition, failed their first objective, that of incident prevention. To ensure an effective incident management, the service operation followed ITIL. Therefore, an analysis was necessary to determine whether the process implementation was effective for all services. Naturally, the efficient use of resources was a condition for successful cost control. The following had to be measured:
 1. Compliance of the ITSM process *Incident Management* with both ITIL and to-be processes.
 2. Maturity of the ITSM process.
 3. Efficiency of the ITSM process for all IT services.
- 2. Define what can be measured.** Three processes were available: the ITIL reference model; a to-be model that was adapted from ITIL; the incident management process for each of the services e-mail, S_1 , S_2 , and S_3 that was implemented in a WFMS based on the to-be model. The WFMS provided a complete history of action for each incident and aggregated all process activities that were done by one employee in one workflow step. The history of action was therefore documented on a higher aggregation level than the to-be model. The calculation of the defined key goal indicators and key performance indicators was also done based on the WFMS data. Therefore, the results of the compliance analysis and the key indicators can be considered consistent with the base data. The use of the data from the WFMS also ensured the reproducibility of findings and facilitated the comparability of data of different periods of time including before and after changes of the as-is process.

5. Business Process Control Framework

The above three pieces of information – model M_1 , model M_2 , and the history of action – led to the measurement of the following:

1. The compliance with specification was measured by comparing two given models. Comparison (M_1, M_2) provided insight as to the degree of to-be model's compliance with ITIL. Comparison (M_1, M_3) analyzed the process maturity of the WFMS. Comparison (M_2, M_3) measured the compliance of the current incident management process with the to-be model.
2. From the results of the compliance analysis, the as-is condition was derived and subsequently be juxtaposed to the to-be condition to determine the process's maturity. Action alternatives for the improvement of ITSM processes could be deduced from the discrepancies.
3. The efficiency was measured by comparing any two given IT services as part of an internal benchmark. The as-is processes and metrics of performance were derived from the histories of action. The processes and the performance values were then contrasted.

3. Gather the data. The eEPC model M_1 was manually derived from the conceptual guidelines of ITIL's incident management. Activities include *Create incident*, *Categorize incident*, and *Prioritize incident*. Based on the process's available documents, the eEPC model M_2 (see Figure 5.17) was manually created. It was aligned with the organization's strategy and goals.

Incidents from a large set of the ISP's IT services were selected from the history of action. The selection comprised only those that were closed within a specified period of time and belonged to the services e-mail, S_1 , S_2 , and S_3 for the automatic extraction of as-is process models to be derived by process mining. It was possible to extract the incidents required for the analysis of these models from the WFMS in an automated way.

4. Process the data. For the goals of measuring the compliance and process maturity, model M_3 ⁶⁹ was derived as described in Section 4.2.4. Prior to the compliance analysis, the process models M_1 , M_2 , and M_3 were converted to Petri nets. The mapping among events was manually performed because events were still present in different granularities.

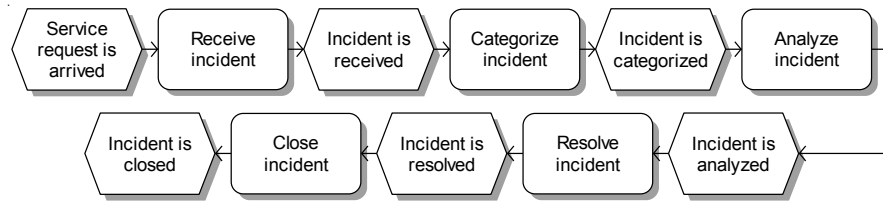


Figure 5.17.: To-be process model Tel-1

⁶⁹ Service e-mail exemplifies the processing of the data of use cases Tel*.

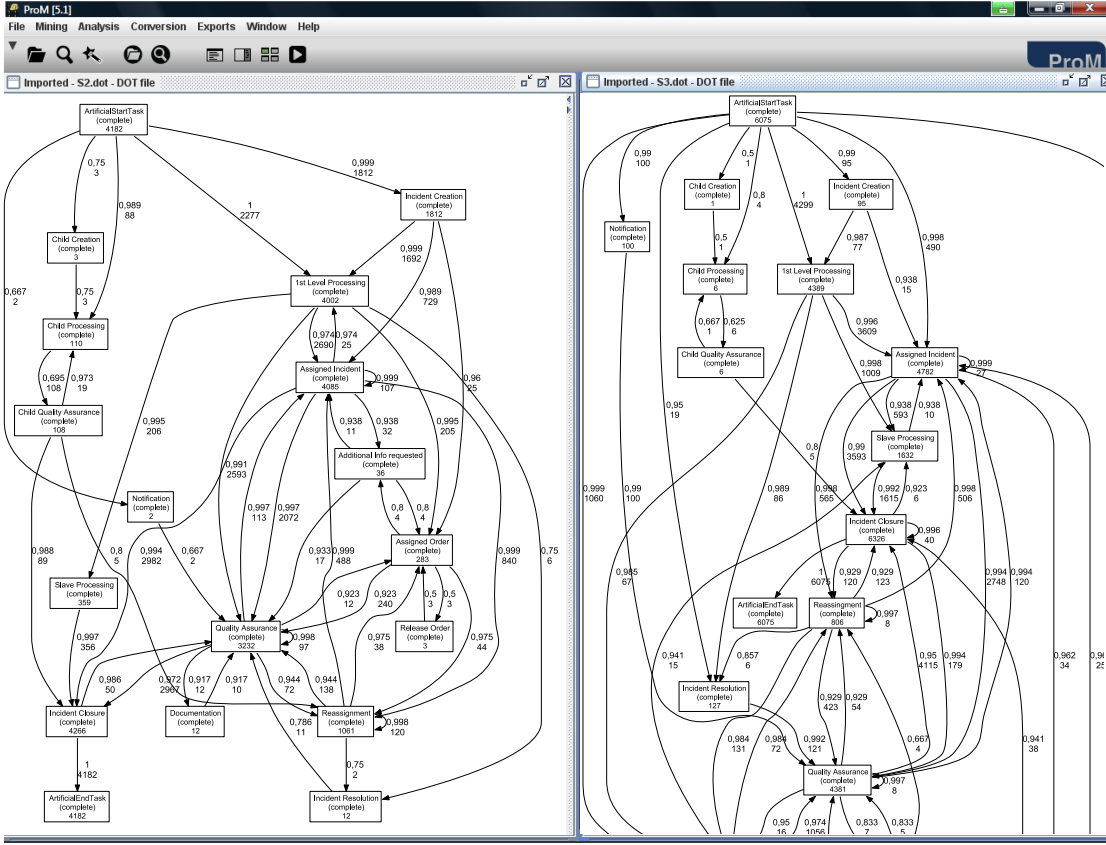


Figure 5.18.: Process models Tel-2 and Tel-3

The data processing that was necessary for the efficiency analysis of limited resources merits attention. Like model M_3 , the as-is process models of the IT services S_1 , S_2 , and S_3 (i.e., M_{3-S_1} , M_{3-S_2} , and M_{3-S_3}) were derived as described in Section 4.2.4. The left-hand side of Figure 5.18⁷⁰ shows model M_{3-S_2} , the right-hand side shows model M_{3-S_3} . The models immediately appear to be different, but close inspection manifests these similarities: same activities, similar starting activities, and similar routing.

The event log further served as an input for a table where all process activities and their absolute (#) and relative (%) occurrences were listed. This listing of the information from M_{3-S_1} , M_{3-S_2} , and M_{3-S_3} made a statistical analysis of the data possible. The functions' mean and standard deviation (SD) provided the statistical relevance. Table 5.1 and Table 5.2 show the indicator-based measurement system. The former contains an excerpt of the complete quality indicator list – the latter highlights indicator candidates for inefficiency.

⁷⁰ See Appendix C for all models M_{3-S_1} , M_{3-S_2} , and M_{3-S_3} .

5. Business Process Control Framework

By gradually narrowing this list of information (i.e., Table 5.1) down to a specific sample of key performance indicators, it, the list, became more comprehensible. This narrowing – a selection process – corresponds to the funnel method (see Figure 5.19).

The highlighted values in Table 5.2 represent the indicators strongly suggesting those inefficiencies that needed further inspection. The values highlighted were outside the range that was computed by the function mean \pm SD. As a result of this process, the following key performance indicators were selected: *Assigned Incident*, *Incident Resolution*, and *Reassignment*.

The processing of the data as described in this section was executed for all candidate service differentiators, such as support group, so that a hierarchy of models and corresponding tables of performance indicators were extracted.

- 5. Analyze the data.** The analysis was based on Section 5.4.2-1, that is, on the three things to be measured noted therein.

First, the compliance analysis of the ITSM process incident management with both ITIL and to-be processes was based on models M_3 and M'_3 .

Since model M'_3 corresponded to the activity level, the compliance analysis was able to be carried out automatically. To quantify compliance with fitness, the ProM plug-in *Conformance Checker* was used. Figure 5.20 depicts comparison (M_2, M_3). Take activity *Categorize incident*. It was executed 735 times – it was not executed nine times. In 567 cases, classifications were made though these activities were not planned (-567). The fitness of comparison (M_2, M'_3) was .78, that of (M_1, M'_3) .60, and that of (M_1, M_2) .92. These values manifested a moderate-to-high compliance with each other.

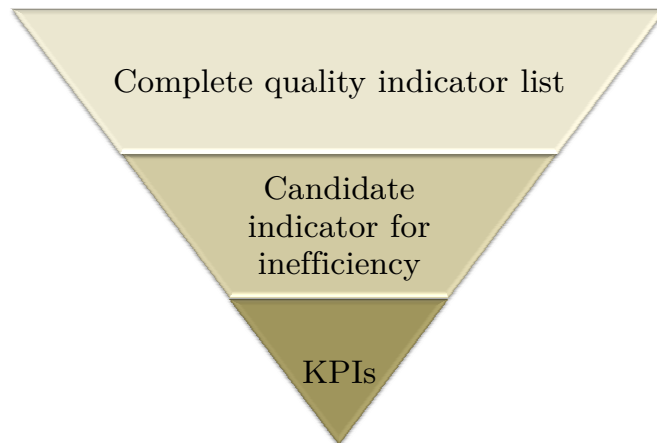


Figure 5.19.: Key performance indicator selection process

Table 5.1.: Excerpt of the indicator-based measurement system

Activity	S ₁		S ₂		S ₃		Mean	SD	Range	
	#	%	#	%	#	%			From	To
Assigned Incident	1,388	76.4	4,085	97.7	4,905	79.4	84.5	11.5	73.0	96.0
Incident Closure	1,815	99.9	4,182	100.0	6,172	99.9	99.9	0.0	99.9	100.0
Incident Resolution	8	0.4	12	0.3	129	2.1	1.0	1.0	-0.1	1.9
Reassignment	238	13.1	1,061	25.4	857	13.9	17.5	6.9	10.6	24.3

Table 5.2.: Indicator candidates for inefficiency

Activity	S ₁		S ₂		S ₃		Mean	SD	Range	
	#	%	#	%	#	%			From	To
Assigned Incident	1,388	76.4	4,085	97.7	4,905	79.4	84.5	11.5	73.0	96.0
Incident Closure	1,815	99.9	4,182	100.0	6,172	99.9	99.9	0.0	99.9	100.0
Incident Resolution	8	0.4	12	0.3	129	2.1	1.0	1.0	-0.1	1.9
Reassignment	238	13.1	1,061	25.4	857	13.9	17.5	6.9	10.6	24.3

5. Business Process Control Framework

The results of the comparison (M_2, M'_3) revealed that the log traces differed from model M_2 in five ways. (Figure 5.21 shows where replay problems have occurred in the log. The activities indicating failures are heavily-shaded.)

(1) Even where the beginning of the incident processing was precisely described, there were log traces, such as T1565085, starting with a different start activity than *Receive incident*. (2) There were log traces, such as log trace T1596654, that included a repetition of activities (i. e., *Analyze incident*). (3) Log traces like T1599057 were observed in which one or more activities (e. g., *Categorize incident*) were missing. (4) Traces were detected in which the right activities were executed in a sequence regardless of the specified order of model M_2 – an illustration is trace T1579962 in which the activity *Resolve incident* was erroneously followed by the activity *Analyze incident*. (5) Traces were found, such as log trace T1601621, that were not properly closed with the designated activity *Close incident*.

To explain these deviations, the fine-grained model M_3 was used. The action flow was assessed manually. Actions such as *Child processing* were extant though they ought to be disused. The intersection of the individual actions provided valuable information about the routing of the incident tickets within the workforce of the ISP. Both structure and frequency of the occurrence of actions were informative. The activity *Reassignment*, for example, designed to reopen erroneously completed incidents, took place quite often, a clear indication that the number of erroneously completed incidents themselves occurred often and there was considerable room for improvement. Models M_3 and M'_3 were both important for the CPSI.

Returning to the log analysis, the investigation revealed two additional measures: the successful execution measure and the proper completion measure. The former expresses the ratio of log traces that were successfully executed to the number of occurrences per trace. The latter expresses the ratio of log traces that were properly completed to the number of occurrences per trace. The successful execution and the proper completion yielded both .18. These low values and the high number of log traces that skipped the required start activity *Receive incident* (620) (see Figure 5.20), led to the conclusion that the start activity indeed

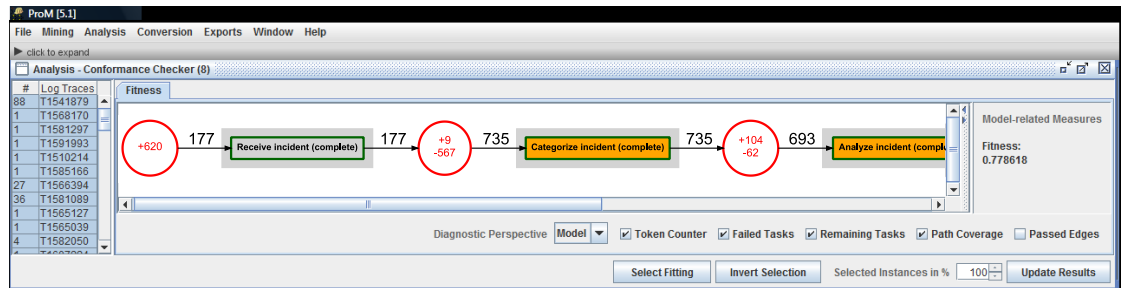


Figure 5.20.: *Conformance Checker* plug-in of model Tel-1

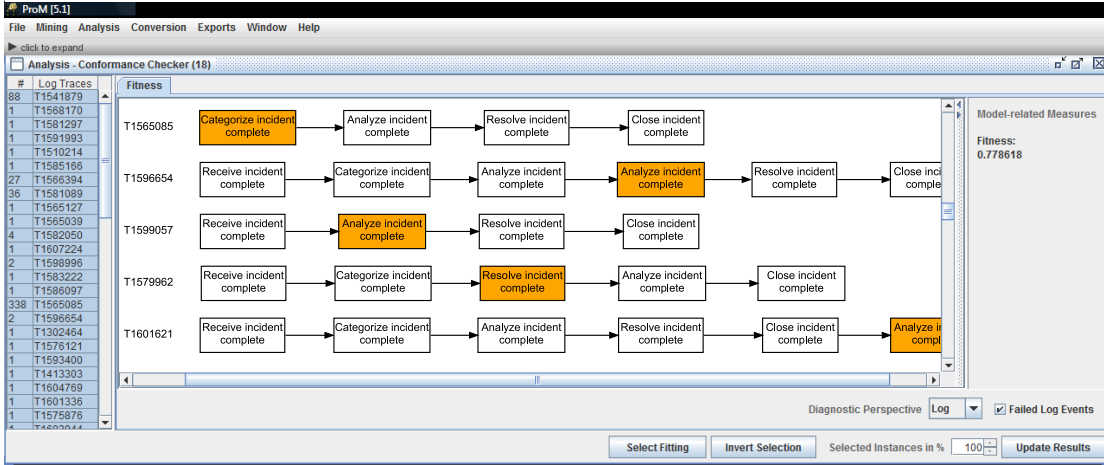
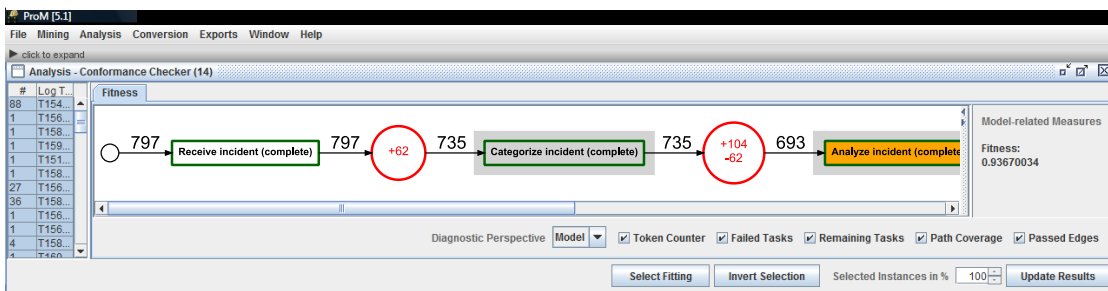


Figure 5.21.: Deviations between log and process model Tel-1

had had a powerful impact on the overall fitness. After having deleted all activities *Receive incident* with the *Event Log Filter*, the required start activity was inserted with the help of the *Artificial Start Task Filter*. Using the *Conformance Checker* again (see Figure 5.22), the fitness value increased to .94 and the number of log traces successfully executed and properly completed both rose to .65. This seemingly modest experiment showed that the ISP could substantially improve the compliance of model M'_3 with model M_2 by ensuring that the start activities were executed as specified.

Studying the effect of rare behavior on the fitness value brought attention to bear on the behavior generally. Indeed, rare behavior's exclusion allowed the analysis of common behavior. Take the instances that occurred least frequently and exclude the rarest 10 % – the fitness increased from .936 to .938; excluding 23 % yielded

Figure 5.22.: *Conformance Checker* plug-in of model Tel-1 showing the result after the adaptation was done

5. Business Process Control Framework

a fitness value of .95. These very small increases stemmed from both the small number of activities of model M_2 and the aggregation of various behavior patterns of model M_3 into the more common patterns of model M'_3 .

The comparison (M_1, M_2) demonstrated that the fitness was very high, meaning that the major objective of designing the ITSM processes according to ITIL was successful.

The comparison (M_1, M'_3) provided a more moderate result. This stemmed both from the adaptation of model M_1 to the business needs of the ISP and from the fact that model M'_3 only reflected information recorded by the WFMS. Manual activities, for example, were excluded from the compliance analysis.

Second, the result of comparison (M'_3, M_2) was incorporated into the maturity model as initial estimates of the as-is situation and was juxtaposed with the to-be situation that was manually determined. The fitness value of comparison (M'_3, M_2) served as a reference value for future comparisons. The compliance analyses were expected to yield higher fitness values if the quality of ITSM processes increased.

Third, the efficiency of the ITSM process for all IT selected services was analyzed based upon inspection of the indicator-based measurement system. A number of performance values deviated substantially from those of different services. Three key performance indicators were discussed in detail: Assigned Incident, Incident Resolution, and Reassignment.

Take the beginning of the incident processing, which is closely connected with the indicator *Assigned Incident*. The indicator showed that the service desk staff were unable to resolve the operational problem themselves and therefore assigned the incident to the next appropriate level. The slight disparity of the frequency between S_3 and S_1 (79.4 % to 76.4 %) indicated the similarity of the two services. Unlike S_1 and S_3 , S_2 's activity was observed to be 97.7%, exceeding the 1.14 SD. Note that the indicator *Assigned Incident* influenced the KPI first kill rate, which was measured with the traditional system of the ISP. As the number of incidents released by the service desk increased, so did the first kill rate. Bear in mind, that the ISP was properly not striving in every case for the highest possible first kill rate, and that the probability of this increased with a given service complexity. For example, it may have been prohibitively expensive to instill the requisite expertise into the part of employees. As in the case for service S_2 , the ISP accepted a lower first kill rate. DEV-1 is the name of this type of deviation.

The second indicator *Incident Resolution* represented the resolution of incidents in which a third party was involved. In services S_1 and S_2 , the activity was present with relatively low frequencies (4 %, 3 %) – in contrast to 2.1 % in S_3 . Statistically speaking, S_3 differed from both S_1 and S_2 by approximately 1.15 SD. The underlying collaboration in S_3 was examined to better understand this deviation's meaning. As noted above a telecommunications carrier was involved in the delivery of service S_3 . Since the resolution of incidents was much more complex,

the deviation (henceforth “DEV-2”) very likely had its origin in the complexity of service S_3 .

The activity *Reassignment* was designed to redirect incorrectly assigned incidents. According to specifications, this activity ought to have resulted only from exceptions. It is noteworthy that the execution of this activity in S_2 exceeded the 1.15 SD. To understand this deviation (henceforth “DEV-3”), the process models were drilled down to the group specific models, and the corresponding table⁷¹ revealed that the working procedure anomalies could be traced to a handful of support groups, a fact strongly suggesting that this deviation resulted from limited resources, quite possibly inadequate knowledge.

In sum, services S_2 and S_3 were particularly deviant, meaning that service S_1 was evidently the most mature service. Among the other services deviations were identified, DEV-1 and DEV-2, that were either intrinsic to the service or was a deviation, DEV-3, stemming from improvable resources. The services observed were distinct in terms of resources and complexity. The factor resource in DEV-1 and DEV-3 differed – the former is a sub-optimum (i.e., low first kill rate) that the ISP accepted when it entertained the entire service, and the latter needed to be improved to optimize the service delivery.

6. Present and use the information. The necessary actions to improve both the ITSM processes and the service delivery were determined in a series of workshops within the organization, the workshops addressing the standardization in the field of incident processing, integration of process mining, control of IT services, and assurance of business processes’ compliance. In view of the increased transparency of service-specific characteristics, the ISP considered the use of the CPSI approach in further process domains.

7. Implement corrective actions. Four corrective actions were taken to address the weaknesses. First, the WFMS was adapted and communicated to those affected by the changes. Second, the WFMS users’ attitude toward compliance was taken into account because WFMS’s users occasionally circumvented licit working methods. Say that the initiation of a request for change (RfC) via the incident management process was disabled in the WFMS – in response, users were trained. This training included both a review of standard procedures for handling changes and the importance of compliance with the change management process. Third, supplementary training for the users closed the knowledge gap. Finally, the comparison base of process mining was adjusted since the ISP accepted the service-specific characteristics in the process.

Additionally, the process manager of the ISP verified both that the use of the CPSI approach clearly enabled a growing maturity of the ITSM process according to the COBIT maturity model and that service delivery was optimized. This growing maturity was manifested by various maturity attributes, such as *tools & automation*

⁷¹ In contrast to Table 5.2, this table involves support groups rather than process activities.

5. Business Process Control Framework

and *awareness & communication*. As to the maturity attribute *tools & automation*, the ITSM processes achieved a higher level of maturity due to the CPSI approach. The approach allowed for both enhanced tool support (maturity level four) and enabled the automation of several resource-intensive parts of the improvement process (maturity level five). The automation was enabled by the first phase of the CPSI approach in which steps three to seven were automatically performed. The presentation of the as-is process models M_3 and M'_3 in step six strongly supports the maturity attribute *awareness & communication*.

5.5. Conclusion

Practical cases from the airline industry (Gerke and Tamm, 2009a) and telecommunications sector (Gerke et al., 2010a) have enabled the development of a procedure model to improve continually both business processes and ITSM processes with process mining and ITIL. And these business examples have proven the approaches propriety. Particularly, it has been shown that the CPSI approach provides guidance not only to comply with reference models but to identify and correct service-specific weaknesses of the process implementation, thereby integrating processes, personnel, systems, and resources.

Two examples for service-specific weaknesses follow. Services of widely different complexity often demand different knowledge levels. Depending on the complexity of the collaboration mode, working procedures can diverge within a process.

The ISP and the airline have attained transparency of their current business processes and ITSM processes, which were continually evaluated based on quality indicators and sorted by degree of maturity. The capability of process mining to reveal hidden information has been particularly useful for dynamically suggesting performance indicators pointing to potential efficiency problems. It has to be stressed that the selection of the indicators is responsive to business objectives that are reflected in varying process executions, indeed a dynamic process. These indicators contribute to an optimization of the IT service delivery in the users' eyes.

Quality degrading and quality improving factors have been determined. The two organizations have used compliance and performance as indicators pertaining to relevance, applicability and practicability of ITIL reference processes. Measures for the CPSI have contributed to an optimization of business process and IT service quality as perceived by the user.

A number of potential benefits have been identified. First, processes and service-specific characteristics of the incident management process have been transparent. Second, the process quality has proven to be a thing that can be measured and controlled through quantifiable information. Third, the degree of applied reference processes has been measurable. Fourth, measurement has been reproducible, repeatable, and comparable, meaning there has been a basis for improvement measures and the corresponding post-control. Last but not least, the maturity level with reference to the COBIT process maturity model has shown itself to be improvable.

Process mining has proven to be an important management tool for re-engineering efforts not only in process-aware processes. However, the difficulties of process mining noted in Section 4.3 reared their heads as to the broader use of the approach. Naturally, these difficulties must be grappled with about which more in due course.

In the future, the approach will be applied both to more complex IT service operation processes (e.g., problem management) and to processes of other life cycle steps (e.g., change management of life cycle *Service Transition*). It will also be interesting to determine if it is possible to apply the CPSI approach to arbitrary IT processes (e.g., software development process) when applying the approach to other business and industry domains. A to-be model consonant with service peculiarities needs to be derived – such a model would serve as the one against which the as-is process could be checked. Furthermore, methods must be developed to build a knowledge base as an input for pattern analysis – information about prior deviations, such as solution, type, and reason, could flow in such a base and the pattern analysis could automatically classify deviations and simultaneously present suggestions to solve them. Finally, traceability ought to be incorporated into the CPSI approach, enabling organizations to identify quickly problematic aspects of their running processes.

6. Measuring Process Compliance with Reference Models

The previous chapters have shown that process mining has considerable potential for assessing the compliance with reference models. Nevertheless, results of the compliance analyses herein using process mining and equivalence algorithms have not yielded results that conform with a given organization's perception of its actual performance. Not every organization could fathom its low compliance values. The differences between the compliance result and the perception have been due in large major to different levels of details, partial view of process mining, and overemphasis on the order of activities. The level of detail characterizing a process tends to differ widely when comparing a reference model with an as-is or to-be process model. The derived as-is model often only partially represents the processes of a given organization. The execution of the processes does not only result in event logs but in written record files, manual activities, and human knowledge. Information outside the reach of process mining algorithms may compromise the results of compliance. Reference models typically do not state whether dependencies among activities are compulsory. As seen in Section 5.4 as to compliance using current equivalence algorithms, the order of activities strongly distorted the compliance result. However, in the case where a reference model does not compel a specific order for the execution of activities, this ordering is not expected to produce distortion. This chapter⁷² therefore takes as its subject process compliance with reference models.

Section 6.1 reviews the related literature. Section 6.2 establishes the importance of measuring the compliance of process models with reference models. Section 6.3 discusses current process quality indicators as to compliance and posits for the need of an algorithm to measure the compliance with reference models. Section 6.4 presents the concept on which this new algorithm is developed especially vis-a-vis overcoming these quality indicators' drawbacks. Section 6.5 measures the compliance of the use cases with reference models in practice. For the purpose of validation, the compliance results are juxtaposed with two extant approaches to explain why current algorithms are not suitable for the evaluation of processes' compliance with reference models. Based on the findings, Section 6.6 concludes.

6.1. Related Work

This chapter presents an approach for measuring the compliance of processes with reference models. Accordingly, other approaches are appraised that also target process

⁷² This work is published in Gerke et al. (2009a) and Gerke et al. (2009b).

measurement and that can be related to compliance or reference models, or to both of them. Quantifying compliance with reference models entails analyzing the commonalities and differences of the processes. An exhaustive analysis of the literature showed a number of distinct approaches, each of which comes under the one of the following: notions of equivalence, notions of similarity, and string matching.

6.1.1. Equivalence Notions

Classical equivalence notions determine whether processes are equivalent in the dynamics of their representation, that is, the process model. There are two classes of notions: the *linear time* and the *branching time*. Trace equivalence is the classical example of linear time equivalence as described in van Glabbeek and Weijland (1996); observation equivalence and bisimulation equivalence, introduced by Milner (1971), are the standard examples of branching time equivalence. The work of Pomello et al. (1992) neatly illustrates equivalence notions for concurrent systems in the framework of Petri nets. The weakest notion is trace equivalence – it determines a process by its possible sets of executions (instances). Being trace equivalent means having sets of identical executions. In contrast, bisimulation and various kinds of observation equivalence consider in addition the branching structure of processes, meaning that if the first process executes a step, then the second process should execute this step. Consider the identical sets of executions $\{ABC_1, ABC_2\}$. The processes are equivalent under trace equivalence but might be different in the case that the choice for C is made in one process after executing A and the same choice is made in another process only after executing B. Given the fact that bisimulation accounts for the point in time at which the decision is made, it is a stronger notion than trace equivalence. The notion of bisimulation is sometimes refined by adding additional requirements and constraints. For example, branching bisimulation as employed in van Glabbeek and Weijland (1996), attempts to capture the moment of choice even if the Petri net includes a notion of silent behavior, namely invisible transitions. The comparison of equivalences of processes with silent behavior described by van Glabbeek and Weijland might be interesting in this respect. Based on equivalence notions, other relations have been introduced.

For example, Basten and van der Aalst (2001) introduced the relations of behavioral inheritance that can also be used to identify commonalities and differences in process models. The approach improves the reusability and the adaptability of process models and concentrates on applying the idea of inheritance derived from object-oriented modeling. Inheritance relations are based on labeled transition systems and use branching bisimilarity in tandem with the algebraic principles of encapsulation and abstraction, the former corresponding to blocking activities, the latter corresponding to hiding activities. Based on inheritance relations van der Aalst and Basten (2001) developed the two inheritance concepts greatest common divisor (GCD) and least common multiple (LCM) of two process models to find a common superior subclass of the process models. The GCD is used to deduce commonalities for a given set of similar process models, while the LCM allows for the construction of the smallest component.

The inheritance concepts based on branching bisimilarity are the source of the delta analysis as proposed by van der Aalst (2003, 2005). This analysis compares the real behavior of an IS with the expected behavior (e.g., a reference model) for the purpose of business alignment.

The objection to the classical equivalence notions including process inheritance against the measurement of process compliance, is that they are defined as a verification property. They return a simple true/false statement (i.e., two processes are equivalent or not) only, failing to provide the degree of equivalence (van Dongen et al., 2008). This is woefully inadequate. A strong argument against equivalence notions is that one presumes that virtually all organizations implement only parts of the reference model. And these notions suffer too by virtue of the fact that a reference model needs to be adjusted to organizational needs – in response to these adjustments the model of course deviates from the common source. Therefore, an equivalence analysis of a process model with a reference model based on classical equivalence notions will very likely fail to be compliant.

6.1.2. Similarity Notions

There are also notions that seek the similarity of processes. Since they assume differences, they naturally determine the similarity between process models rather than the equivalence.

Notions of similarity⁷³, such as fitness and precision, describing a process model's completeness and appropriateness are originally used to verify the correctness of a process model. Given the fact that it is possible to verify whether or not an event log or a process model corresponds to a process model, measuring the compliance can be deemed a specific form of process verification. One idea of model verification is both to replay the event log's instances from which the process model originated in this very model and to determine if this model holds true for the instances. This approach can be used not only to verify the quality of the model but also to measure the correspondence of a given event log with respect to a predefined process model, notably a reference model, specifying how people and organizations are expected to work. Van der Aalst introduced the concept and this approach is referred to as conformance testing for the purpose of business alignment. The *Conformance Checker* plug-in available in ProM is an example for an implementation of conformance testing. The plugin quantifies the fitness, the precision, and the advanced structural appropriateness of a process model with respect to a given event log, and it has been employed in various contexts. Measuring the conformance of the behavior of web services is such an employment (van der Aalst et al., 2008). Conformances' pertinence in the context of security violations has been investigated as well. Van der Aalst and Alves de Medeiros (2005) related security events (e.g., provide a password before processing an order) to the instances of a process model. Since security typically applies only to a part of the process, they checked the conformance for the tasks involved in a subnet of the process model. The cited authors therefore inspected

⁷³ They will be explained in Section 6.3.

6. Measuring Process Compliance with Reference Models

the event logs for the presence or absence of certain patterns to a substructure of the process model.

Extensions of equivalence notions are used to search for similar behavior. Dijkman (2008), for example, categorized differences related to control flow, resource assignment, and activity correspondence and presented a technique to ascertain these differences based on the notion of complete trace equivalence. Ehrig et al. (2007) presented an approach for measuring similarity between semantic business process models that are an instantiation of the predicate/transition net ontology.

6.1.3. String Matching

In the area of string matching, it is common practice to distinguish between the exact string matching and the approximate string matching. The string matching is the approach of locating a pattern string as a substring of a text. Hume and Sunday (1991) and Pirklbauer (1992), among others, described major categories and characteristics of exact string matching algorithms. Searching for a string in a text is undeniably a valid concern of computer science and its applications (Hume and Sunday, 1991), but these strings are of limited aid for the same reasons classic equivalence notions are often unhelpful. The approximate matching approach concerns subsequences. Say the string *xyz* is a subsequence, but not a substring in *axayaz*. The shift from substrings to subsequences is a shift from exact matching to approximate matching – an approach that seeks to locate a pattern string in a text when the differences between the pattern and its occurrence in the text (Navarro, 2001) is limited. From the numerous models that measure similarity between two character strings, the most widely known is the edit distance method as proposed by Levenshtein (1966). The edit distance between two strings *x* and *y*, that is, $ed(x,y)$, is defined by the minimum number of edit operations (addition, deletion, and substitution of characters) necessary to transform the source string *x* to the target string *y*, or vice-versa (Navarro, 2001; Needleman and Wunsch, 1970). For example, the edit distance $ed(validation, verification)$ equals five because three substitutions and two additions are required to transform the string *validation* to the string *verification*. The greater the edit distance, the more vary the strings. The *longest common subsequence* (*lcs*) approach (Needleman and Wunsch, 1970; Bergroth et al., 2000; Navarro, 2001) is a special case of the edit distance problem; it allows only the operations addition and deletion. The *lcs* measures the length of the longest pairing of characters that can be made between two strings so that the pairings are consistent with respect to the order of the letters (Navarro, 2001). The survey of Bergroth et al. (2000) gives a thorough comparison of well-known *lcs* algorithms. The Damerau-Levenshtein distance allows addition, deletion, substitution, and the transposition of two adjacent characters. The Hamming distance allows only substitution and hence, only applies to strings of the same length.

Cook and Wolf (1999) and Cook et al. (2001) introduced a technique, process validation, that compares the event stream coming from the process model with the event stream from the execution log based on two different string distance metrics. Note that this technique's creators did not preclude of assigning weights in order to differentiate the relative importance of specific types of events.

It is possible to match a string to a regular expression. LTL suggested by Pnueli (1977), for example, provides an elementary way to formulate linearly time constraints. Van der Aalst et al. (2005) used them to verify certain properties of a process execution, and with that contribute to compliance. The *LTL Checker* plug-in (de Beer, 2004) available in ProM analyzes if instances satisfy a given property. The properties in question, such as “is activity A done by person P and Q” or “is activity B always executed after activity A is carried out”, are specified based on temporal formulas. They denote qualities a process execution can or cannot have.

In contrast to the previous applications, Juan (2006) employed a string comparison approach as to instances embedded in each process model to identify differences.

In sum, both similarity notions and string matching approaches are used to determine commonalities of process models. The works focusing on differences, illustrations being afforded by Dijkman (2007, 2008) and Juan (2006), do indeed complement this dissertation – they facilitate locating the exact position of a difference from the reference model and specify the type of that difference. Common integration approaches for process models show how these distinctions can be integrated, for example, to harmonize processes after organizations’ merger (Frank and Eder, 1999; Küster et al., 2006; Mendling and Simon, 2006). Since only a small number of researchers have investigated the question of how to measure the compliance with reference models, this topic provides much untiled earth.

6.2. Requirements of Measuring Compliance

Process compliance is defined as the degree to which a process model behaves in accordance with a reference model. This behavior is expressed by the instances that can be generated by the process model.

Figure 6.1 shows two extended event-driven process chains capturing similar functionalities. Both are taken from use case Air-1. The airline has adopted the ITIL reference model M_1 to improve customer relations. A team of experts, the members of which were chosen from among the companies’ employees, studied the characteristics of the given process and thereby devised a to-be process model M_2 . The two process models M_1 and M_2 are different as can be seen in Figure 6.1.

The eEPC in the lower part of the figure shows model M_1 , which depicts two activities represented by these functions: *Create incident* and *Categorize incident*. The eEPC in the upper part of the figure shows model M_2 . Processing starts with a complaint made in an e-mail or by filling out a form on the airline’s website. In the latter case, the customer has to classify the complaint by choosing from given categories (e.g., delay and lost & found). In the former case, an employee has to read the e-mail and to determine the category manually. The processes share two equivalent functions and two corresponding events. The events *Incident record is created* and *Incident is categorized* in model M_1 correspond to the events *Claim is created* and *Claim category is assigned* in model M_2 . Function *Create incident* in model M_1 reflects the corresponding function *Create customer inquiry* in model M_2 . Function *Categorize incident* of models M_1

6. Measuring Process Compliance with Reference Models

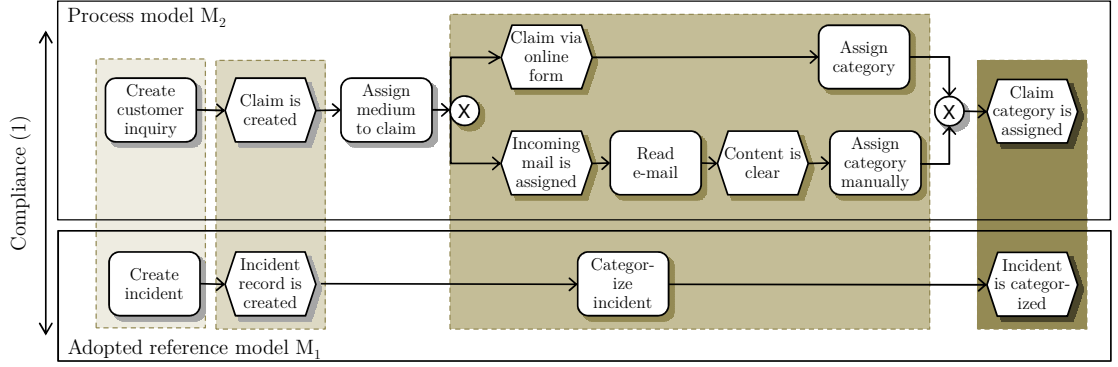


Figure 6.1.: Receipt of complaints at an airline

corresponds to a set of events and functions. The corresponding function and activities are highlighted in model M_1 and model M_2 . To measure compliance, characteristics of business and reference models must be considered.

In real life, an organization may use numerous notations. This is not a limitation since tools, such as ProM, can convert one modeling notation to another. In any event, the assumption here is that compliance will be measured between models that are expressed in the same language. The notation used in the use case is eEPC because it is the general applied modeling notation of the airline.

6.2.1. Compliance Maturity and Degree

The case study has identified two major concerns as to evaluating compliance with reference models. First, the airline wanted to learn if its processes followed the behavior recommended by the reference model. Second, the airline wanted to learn if all the behavior recommended by the reference model was being implemented. In the context of compliance, the first concern is referred to as compliance degree and the second as compliance maturity. Take the processing of customer complaints. Model M_1 recommends accepting complaints either only via e-mail, letter, or phone. When the airline accepts complaints via e-mail or letter, only a part of the recommendations is implemented, making the airline partially mature with respect to compliance maturity. E-mail and letter complaint implementation correspond to model M_1 's recommendations, making the airline fully compliant with respect to compliance degree.

6.2.2. Granularity of Models

Having two models, M_1 and M_2 , it may happen that the granularity characterizing the level of detail of activities varies. It is possible that compliance applies to a set of activities, rather than to individual activities. For example, activity *Categorize incident* of model M_1 corresponds to a set of activities in model M_2 highlighted in Figure 6.1. The correspondence of activities needs to be identified to account for the granularity.

Correspondence is a mapping among activities of model M_2 to activities of model M_1 where the functionality of the activities is the same. Current approaches, such as schema and semantic matching (van Dongen et al., 2008; Ehrig et al., 2007), assume that the correspondence can be established automatically based on the labels of the activities. The examples of the use cases have shown that this assumption is unrealistic. For instance, the activities *Create incident* in model M_1 and *Create customer's complaint* in model M_2 have the same functionality, but they have different labels. Since the automatic mapping is not applicable, the activities herein must be manually mapped.

6.2.3. Customization of the Reference Model

Note that not every part of model M_1 can be treated the same way when measuring compliance. For example, since reference models do not typically state whether activities have to be executed in a specified order, the order may not always be important. These special parts are referred to as a user-selected set of activities (henceforth “user-selected partition”). It has both an identification and a type, the latter being either “Order” or “Exclusion”. Figure 6.2 shows that activities *Categorize incident* and *Prioritize incident* in the user-selected partition \mathcal{P}_1 may be executed in any order – this may be the case, for example, if the prioritization depends only on a specific customer’s characteristics, such as the frequency with which he or she flies. Thus, the activity *Prioritize incident* can be executed without knowing the complaint’s category. A user-selected partition of type *Exclusion* allows the definition of activities of model M_1 that need to be excluded from the compliance analysis. Consider the user-selected partition \mathcal{P}_2 . In the use case, the activity *Preprocess incident* is not supported by the IS right now. Nonetheless, a manual activity corresponding to the functionality expressed by *Preprocess incident* is executed. To prevent the missing activity from erroneously affecting the compliance, the activity is excluded.

6.2.4. Iteration of Process Activities

Consider a special case, one in which an activity is part of an arbitrary cycle (e. g., loop, iteration, and recursion) in process M_3 though it is not in model M_1 . This activity can be executed repetitively, while in model M_1 it must be performed in only one iteration. For example, in the use case, the activities *Search for a solution* and *Inform Customer* are performed repeatedly until the customer accepts the processing of the complaint. Using the cycle shown in Figure 6.2 increases the quality of the process and contributes to a higher degree of customer satisfaction. Thus, even if the reference model does not explicitly recommend a cycle, the airline believes that this cycle in model M_3 does *not* affect the compliance with model M_1 – in contrast with a cycle that merely redoes work. Such a cycle adversely affects the efficiency of a process. Even more confounding virtual all reference models neither contain cycles nor state a precise number of recommended iterations. Generally, knowing little or nothing of the semantics of cycles makes it impossible to assay its effect on compliance.

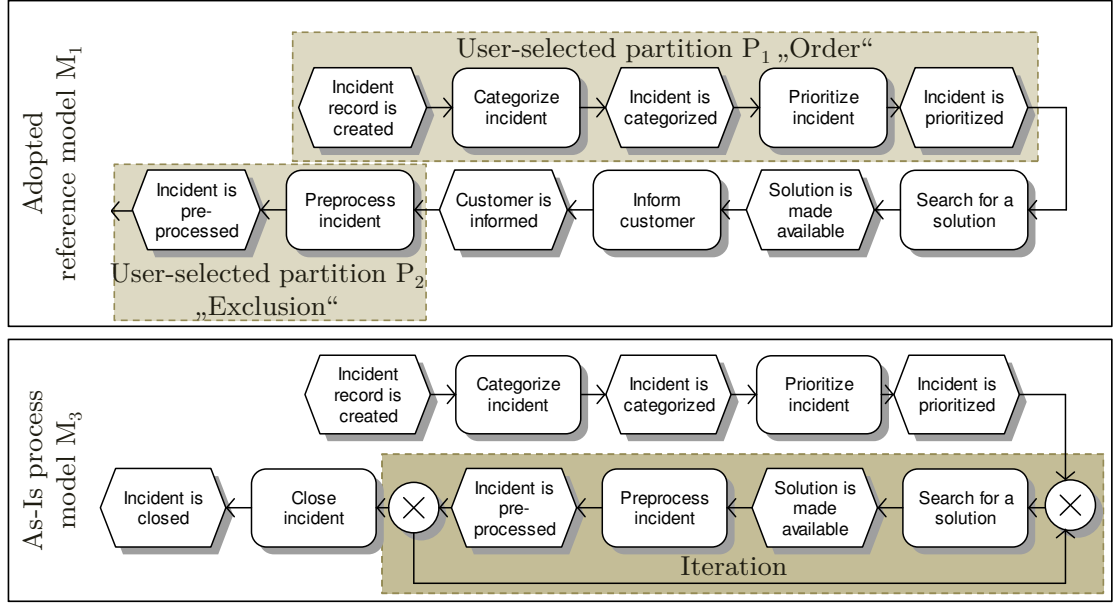


Figure 6.2.: Customization of a reference model

6.3. Measuring Process Compliance with Quality Indicators

As Section 6.1 has shown that classical equivalence notions are not suitable for measuring the compliance of process models with reference models, similarity notions merit close attention.

Table 6.1 depicts well-known quality measurements in the area of process mining, which can be grouped into four dimensions: fitness, precision, structure⁷⁴, and generalization. The first dimension, (D1), fitness (AKA correctness) analyzes the extent to which the instances of an event log can be accurately reproduced in a reference model. The second dimension precision, (D2), relates to the question of the observed behavior's description's accuracy. Since precision or appropriateness measures the ability to refrain from utilizing worthless information, this dimension needs to be balanced with fitness. That is, accuracy needs to be weight against clarity. As to the structure of a process model, (D3), it is determined by the semantics of the modeling language (e. g., AND and XOR semantics), there are several ways to express the same behavior. Obviously, there are preferred ones as well as ill-suitable representations. Generalization, (D4), starts with the fact that a *good* process model's definition has changed slightly in the past two

⁷⁴ The focus herein is on dynamics of process models. Therefore, notions concerned with the syntactical structure of process models, such as place invariants, transition invariants, reduction rules, are not discussed. Interested readers are referred to Cardoso (2007), Reijers and Vanderfeesten (2004), Mendling (2006), and Melcher and Seese (2008).

or three years – elementary and analyzable quality criteria have come to the fore, and with that generalization was introduced. To obtain meaningful models, particularly in the context of more unstructured and flexible processes, Rozinat et al. (2008) sacrificed a degree of fitness and precision for the sake of generalization.

The quality indicators' values depicted in Table 6.1 range from 0 to 1. The closer the value is to 1, the better the process model's quality as to the respective indicator.

Two ways exist to validate the compliance of processes with the supporting IS: log-based analysis and inter-model analysis, each of which depends on a predefined process model, which, for instance, can be extracted by a process mining algorithm.

6.3.1. Log-Based Analysis

The log-based process verification is predicated on the existence of an event log and at least one process model.

Measures to Assess Fitness

Measures that relate to the fitness dimension include *Completeness*, *Continuous Parsing Measure (CPM)*, and *Fitness*⁷⁵, each of which is based on replaying the log in a model. They differ from each other both in what they deem to be the “unit of behavior” and in the underlying model type.

Completeness. It calculates the percentage of the event traces that can be generated by a workflow model (Greco et al., 2006). The event traces express the behavior of the model.

The remaining fitness measures CPM and Fitness consider both traces and activities as their units of behavior, thereby according equal importance to the total number of correctly replayed traces and the total number of correctly replayed activities.

Continuous Parsing Measure. It counts the number of correctly parsed event traces with respect to the total number of event traces in the event log while replaying a Heuristics net (Weijters et al., 2006). Whenever a parsing error occurs, it is registered and the parsing continues. Both missing activities during parsing and hanging activities after parsing negatively influence the CPM.

Fitness. It replays an event log in a Petri net and relates both the number of missing tokens with the amount of consumed ones and the number of remaining tokens with produced ones. Thereby, it punishes both situations in which events cannot be replayed since the corresponding activity is “not activated” and those in which activities “remain activated” (Rozinat and van der Aalst, 2008). Like the CPM, the log replay is carried out in a non-blocking way, meaning that if there are tokens missing to fire the transition in question, they are artificially created, and replay proceeds.

⁷⁵ The *Fitness* is also known as *Token-Based Fitness*.

Table 6.1.: Overview of current verification measures in the context of process mining (from Rozinat et al. (2007a))

Authors	Measures	Dimensions				Input			Output Model Type
		D1	D2	D3	D4	I1	I2	I3	
Greco et al. (2006)	Completeness	✓				✓	✓		Workflow schema
	Soundness		✓						
Weijters et al. (2006)	Continuous Parsing Measure	✓				✓	✓		Heuristics net
	Fitness								
Rozinat and van der Aalst (2006, 2008)	Behavioral Appropriateness		✓			✓	✓		Petri net
	Structural Appropriateness			✓					
	Behavioral Precision		✓			✓	✓	✓	
Alves de Medeiros (2006)	Behavioral Recall				✓				Heuristics net
	Structural Precision & Recall								
	Duplicates Precision & Recall			✓			✓	✓	
Van Dongen et al. (2006)	Footprint Similarity	✓			✓		✓	✓	Petri net, eEPC
Legend: D1 = fitness D2 = precision D3 = structure D4 = generalization									
I1 = event log I2 = mined model I3 = reference model									

Measures to Assess Precision and Generalization

Measures that relate to the precision and generalization dimensions include *Soundness*, *Behavioral Appropriateness*, *Behavioral Precision*, and *Behavioral Recall*.

Soundness. It calculates the percentage of event traces that are both generated by a workflow model and in the event log, albeit with the proviso that the log must contain all possible traces (Greco et al., 2006). When the model has many tasks in parallel the log may not contain all of them, and when the model has loop constructs it cannot contain all of them.

Behavioral Appropriateness. It ⁷⁶ establishes predecessor and successor relations reflecting alternative or parallel behavior among activities in an event log and among activities in a model, and it compares these relations (Rozinat and van der Aalst, 2006, 2008). The more relations (from the model) that can be derived from the log, the more precise is that model.

Behavioral Precision and Recall. It requires as additional input a reference model and measure the intersection of enabled activities that the mined model and the reference model have at every moment of the log replay (Alves de Medeiros, 2006). This intersection is weighted by the frequency of event traces in the event log (Alves de Medeiros, 2006). *Behavioral Precision* measures how much extra behavior the mined model allows for with respect to a given reference model and an event log. *Behavioral Recall* measures the opposite. These measures capture both the moment of choice in the Heuristics nets and the differences of behavior in low and high frequent traces.

Measure to Assess Structure

The *Structural Appropriateness* verifies the quality of a Petri net.

Structural Appropriateness. It assesses the capability of a model to describe the event log in a structurally suitable way. It is based on the detection of redundant invisible tasks and alternative duplicate tasks (Rozinat and van der Aalst, 2006, 2008).

6.3.2. Inter-Model Analysis

The *Structural Precision and Recall*, *Duplicates Precision and Recall*, and *Footprint Similarity* address the inter-model analysis in the context of process mining. Since the two algorithms *Structural Precision and Recall* and *Footprint Similarity* will be used subsequently, their equations are also presented.

Structural Precision and Recall. They equate the term *structure* with all firing sequences that can occur in a Heuristics net (Alves de Medeiros, 2006). Precision is

⁷⁶ Note that this measure refers to the advanced form of the *Behavioral Appropriateness*.

6. Measuring Process Compliance with Reference Models

the fraction of connections in model M_2 that also appear in model M_1 (see Equation 6.1). If this value is 1, the *Structural Precision* is the highest – all connections in the second model exist in the first model. Recall is the fraction of connections in M_1 that also appear in M_2 (see Equation 6.2). If the value is 1, the *Structural Recall* is the highest – all connections in the first model exist in the second model.

$$precision^S(M_1, M_2) = \frac{|C_1 \cap C_2|}{|C_2|} . \quad (6.1)$$

$$recall^S(M_1, M_2) = \frac{|C_1 \cap C_2|}{|C_1|} . \quad (6.2)$$

Duplicates Precision and Recall. They analyze if two Heuristics nets have the same amount of duplicates (Alves de Medeiros, 2006). The *Duplicates Recall* assesses how many duplicates the original model has that are not in the mined model. The *Duplicates Precision* quantifies how many duplicates the mined model has that are not in the original model.

Causal Footprint. It measures the similarity of two models based on their structures – note that either one Petri net is compared with another Petri net or an eEPC is compared with another one. The algorithm *Footprint Similarity* $sim(G_1, G_2)$ operates by mapping the models to their causal closure graphs, transforming these graphs to vectors in space, and measuring the cosine of the angles between those vectors (van Dongen et al., 2006), that is ⁷⁷,

$$sim(G_1, G_2) = \frac{\vec{g}_1 \times \vec{g}_2}{|\vec{g}_1| \cdot |\vec{g}_2|} = \frac{\sum_{j=1}^{|\Theta|} g_{1,j} \cdot g_{2,j}}{\sqrt{\sum_{j=1}^{|\Theta|} g_{1,j}^2} \cdot \sqrt{\sum_{j=1}^{|\Theta|} g_{2,j}^2}} . \quad (6.3)$$

Behavior captured by the log-based measures is always with reference to a given event log, though, naturally, such logs are not always extant. And even if one is available, when it is compared to a model flawed equivalence quantification might occur. Examples include those in which the model does not fit the log, the model consists of activities that are not in the log, and the log does not contain enough information, relative to the actual amount of behavior ⁷⁸ to measure the compliance maturity. This last may occur quite frequently since event logs represent, in the great majority of cases, behavior over a period of time. If a log file is available and it represents the behavior of the organization, a process model can be derived by process mining from that log.

⁷⁷ This formula is complex and therefore it is not fully described. Van Dongen et al. (2008) provide an elaboration.

⁷⁸ As time goes to infinity, the log may virtually fully represent the model.

This dissertation examines the inter-model analysis for the purpose of measuring compliance between two models, bearing in mind that a process model is typical behavior at (re)design time, but an event log reflects the behavior typical at run-time.

6.4. Development of a Sequence-Based Compliance Analysis

Based on the requirements from Section 6.2, an algorithm called *Sequence-Based Compliance Analysis* is developed to measure the compliance of model M_2 , or that of M_3 , with model M_1 . The algorithm can deal with two models having different structures because it can judge whether or not one process is compliant with the other. The example in Figure 6.5 illustrates that the process models are different, but, in due course, it will be shown that they are compliant.

Section 6.4.1 presents theoretical foundations needed for the development of the algorithm in Section 6.4.2. Section 6.4.3 uses a case study showing the algorithm's feasibility in practice. This study is performed by applying the algorithm to a subset of 252 extended event-driven process chains of the SAP reference model. Section 6.4.4 introduces a running example, which is used to illustrate concept and implementation.

6.4.1. Theoretical Foundations

Previous sections have used the eEPC language to model processes since it is lucid and is broadly used in industry, and it is indeed the one used by the airline. A more formal language, one based on workflow nets (van der Aalst, 1997), is used for the design of the compliance algorithm. Its formality is well-suited to analyze processes. The degree of compliance is defined based on the firing sequences of workflow nets.

Definition 9 (Workflow Net)

A *WF-net* is a tuple $M = (P, T, F, i, o)$ so that:

- P is a finite set of *places*,
- T is a finite set of *transitions*,
- $P \cap T = \emptyset$,
- $F \subseteq (P \times T) \cup (T \times P)$ is a set of *arcs*,
- $i \in P$ is the unique *source* place so that $\bullet i = \emptyset$,
- $o \in P$ is the unique *sink* place so that $o \bullet = \emptyset$,
- Every node $x \in P \cup T$ is on a path from i to o , where for each node $x \in P \cup T$ the set $\bullet x = \{y \mid (y, x) \in F\}$ is the *preset* of x and $x \bullet = \{y \mid (x, y) \in F\}$ is the *postset* of x . *

6. Measuring Process Compliance with Reference Models

Transitions represent the activities of an instance. The input place (i) and the output place (o) of the WF-net express the entry point when instances are created and the exit point when instances are deleted. The last requirement of Definition 9 ensures that there are no transitions and places that do not contribute to processing.

Definition 10 (Firing Sequence)

Let $M = (P, T, F, i, o)$ be a WF-net and let $t \in T$ be a transition of M .

- A *marking* $K : P \rightarrow \mathbb{N}$ is a mapping defining the number of tokens per place.
- t is *enabled* in a marking K if $(\forall p \in \bullet t) K(p) \geq 1$.
- t *fires from* marking K to marking K' , denoted by $K[t]K'$, if t is enabled in K , and $(\forall p \in \bullet t) K'(p) = K(p) - 1$, and $(\forall p \in t\bullet) K'(p) = K(p) + 1$.
- $\sigma = \langle t_1, t_2, \dots, t_n \rangle \in T^*$ is a *firing sequence* leading from a marking K_1 to a marking K_{n+1} , denoted by $K_1[\sigma]K_{n+1}$, if there are markings K_2, \dots, K_n so that $K_1[t_1]K_2[t_2] \dots K_n[t_n]K_{n+1}$. *

Example 11 (Firing Sequence)

The model M_1 from Figure 6.2 has only one firing sequence: $\sigma = \langle \text{Categorize incident}, \text{Prioritize incident}, \text{Search for a solution}, \text{Inform customer}, \text{Preprocess incident} \rangle$. ◇

To capture relevant behavior only firing sequences are considered that are terminated properly.

Definition 11 (Complete Sound Firing Sequences)

Let $M = (P, T, F, i, o)$ be a WF-net and $\sigma \in T^*$.

- K_i is the *initial marking* with $K_i(i) = 1$ and $(\forall p \neq i) K_i(p) = 0$.
- K_o is the *final marking* with $K_o(o) = 1$ and $(\forall p \neq o) K_o(p) = 0$.
- σ is a *complete sound firing sequence*, if $K_i[\sigma]K_o$.
- $S(M)$ denotes the set of all complete sound firing sequences. *

This definition deliberately ignores unsound behavior, for example, instances running into a deadlock or a livelock. A marking is dead if it is not a final marking and if no other marking can be reached from it. A marking is live if it enters a cycle that cannot be left. σ denotes a firing sequence.

Since workflow nets can be considered as directed graphs in which $P \cup T$ is the set of nodes and F is the set of arcs, the standard graph-theoretical notion of a cycle is used.

Definition 12 (Cycle)

A *cycle* in a WF-net $M = (P, T, F, i, o)$ is a sequence of nodes $(x_1, \dots, x_n) \in (P \cup T)^*$ so that $(\forall 1 \leq i < n) (x_i, x_{i+1}) \in F$ and $x_1 = x_n$. *

The existence of cycles causes the set $S(M)$ to be in general infinite. Therefore the number of unroll factors for cycles is restricted by a variable parameter⁷⁹. The finite subset of $S(M)$ is denoted by $S'(M)$. The set $S'(M)$ grows exponentially in the number of transitions $|T|$, a seeming indication that the approach would not be feasible in actual business. However, Section 6.4.3 will show that this approach can be used in practice. To deal with cycles and their contribution to compliance among competing requirements (see Section 6.2.4), one equates cycles having no correspondence in model M_1 with the action of redoing work. This redundant work may have a negative effect on the compliance values.

6.4.2. Measuring Compliance

To account for the special characteristics of compliance with reference models identified in Section 6.2, the algorithm has several parameters.

Definition 13 (Granularity Mapping)

Let $M_1 = (P_1, T_1, P_1, i_1, o_1)$ and $M_2 = (P_2, T_2, P_2, i_2, o_2)$ be two workflow nets, where M_1 refers to the *reference model* and M_2 to the *process model*. The mapping $\mathcal{G} : T_2 \rightarrow T_1$ relates activity labels in the process model to activity labels in the reference model. Given the fact that \mathcal{G} can be non-injective, this mapping handles granularity differences between the two models. The term *granularity mapping* is used for \mathcal{G} . *

Definition 14 (User-Selected Partition)

Let M_1 be a reference model as stated in Definition 13. A *user-selected partition* of M_1 is a set of transitions $p \subseteq T_1$, which are of type either *exclusion* or *order*. User-selected partitions of type exclusion are represented with \bar{p} and those of type order with \check{p} . M_1 can have associated with it at most one user-selected partition of type exclusion and a finite number of user-selected partitions of type order. \mathcal{P} denotes the set of all user-selected partitions associated with M_1 . *

Example 12 (User-Selected Partition)

The adopted reference model M_1 from Figure 6.2 describes two user-selected partitions: \mathcal{P}_1 and \mathcal{P}_2 . The user-selected partition \mathcal{P}_1 is of type order and is comprised of the transitions *Categorize incident* and *Prioritize incident*. The user-selected partition \mathcal{P}_2 is of type exclusion and is comprised of the transition *Preprocess incident*. ◇

Having defined the parameters, the compliance measures can be deduced.

Definition 15 (Extended and Mapped Firing Sequence Set)

Let M_1 and M_2 be the reference model and the process model as stated in Definition 13. Let \mathcal{P} be the set of all user-selected partitions related to M_1 and let \mathcal{G} be the granularity mapping between M_1 and M_2 . Let $\sigma_1 \in T_1^*$ and $\sigma_2 \in T_2^*$.

⁷⁹ Note that the parameter is omitted here and in subsequent equations for clarity because it has no significant effect on the equations.

6. Measuring Process Compliance with Reference Models

- $\sigma_1^{\text{ext}}(\mathcal{P})$ is the *set of extended firing sequences* of σ_1 that is derived from σ_1 by applying two actions to σ_1 : (1) remove the transitions in \bar{p} from σ_1 and (2) generate the permutations of $\sigma_1 \setminus \bar{p}$ for all user-selected partitions \check{p} .
- $|\sigma_1|_{\text{ext}} = |\sigma'_1|$ ($\sigma'_1 \in \sigma_1^{\text{ext}}(\mathcal{P})$)⁸⁰ denotes the length of an extended firing sequence σ'_1 of σ_1 .
- $\sigma_2^{\text{map}}(\mathcal{G})$ is the *set of mapped firing sequences* of σ_2 that is derived from σ_2 by applying \mathcal{G} to all transitions of σ_2 , whereas for each subsequence of transitions of σ_2 that is mapped to the same transition $t_1 \in T_1$ only one occurrence of t_1 is placed in the resulting sequences – it may be placed at different positions resulting in several mapped sequences.
- $|\sigma_2|_{\text{map}} = |\sigma'_2|$ ($\sigma'_2 \in \sigma_2^{\text{map}}(\mathcal{G})$) denotes the length of a mapped firing sequence σ'_2 of σ_2 . *

Removing transitions by \bar{p} means $|\sigma_1|_{\text{ext}} \leq |\sigma_1|$ and the mapping of possible multiple transitions to one transition means $|\sigma_2|_{\text{map}} \leq |\sigma_2|$.

Definition 16 (Compliance Measures)

Let M_1 , M_2 , \mathcal{G} , and \mathcal{P} as stated in the definitions above. Let $\sigma_1 \in T_1^*$ and $\sigma_2 \in T_2^*$.

- The *firing sequence compliance* (*fsc*) of σ_2 w.r.t. σ_1 is:

$$\text{fsc}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G}) = \max\{ \text{lcs}(s, s') \mid s \in \sigma_1^{\text{ext}}(\mathcal{P}), s' \in \sigma_2^{\text{map}}(\mathcal{G}) \} . \quad (6.4)$$

- The *firing sequence compliance degree* (*fscd*) of σ_2 w.r.t. σ_1 is:

$$\text{fscd}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G}) = \frac{\text{fsc}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G})}{|\sigma_2|_{\text{map}}} . \quad (6.5)$$

- The *firing sequence compliance maturity* (*fscm*) of σ_2 w.r.t. σ_1 is:

$$\text{fscm}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G}) = \frac{\text{fsc}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G})}{|\sigma_1|_{\text{ext}}} . \quad (6.6)$$

- The *compliance degree* (*cd*) of M_2 w.r.t. M_1 is given by:

$$\text{cd}(M_2, M_1, \mathcal{P}, \mathcal{G}) = \frac{\sum_{\sigma_2 \in S'(M_2)} \max_{\sigma_1 \in S'(M_1)} \{ \text{fscd}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G}) \}}{|S'(M_2)|} . \quad (6.7)$$

- The *compliance maturity* (*cm*) of M_2 w.r.t. M_1 is given by:

$$\text{cm}(M_2, M_1, \mathcal{P}, \mathcal{G}) = \frac{\sum_{\sigma_1 \in S'(M_1)} \max_{\sigma_2 \in S'(M_2)} \{ \text{fscm}(\sigma_2, \sigma_1, \mathcal{P}, \mathcal{G}) \}}{|S'(M_1)|} . \quad (6.8)$$

⁸⁰ Note, that $|\sigma_1|_{\text{ext}}$ is well-defined. The length of each extended sequence $\sigma'_1 \in \sigma_1^{\text{ext}}(\mathcal{P})$ is equal since each differs only in the order of transitions. The same holds for $|\sigma_2|_{\text{map}}$.

Function lcs in Equation 6.4 is an adaptation of the lcs function (Bergroth et al., 2000), which operates on strings. It calculates the length of the longest common subsequence of two firing sequences, thereby finding the maximum number of identical activities while preserving the activity order. The greater the value returns, the more the firing sequences resemble each other. (See Bergroth et al. (2000) for details on lcs .) Since the firing sequences σ_1 and σ_2 can have various structures manifested in their extended and mapped firing sequence sets, the variation of σ_1 and σ_2 will be selected yielding a greater similarity of σ_1 and σ_2 . The compliance degree (see Equation 6.5) of firing sequence σ_2 indicates the extent to which the transitions of σ_2 are executed according to the specifications of a reference model expressed with σ_1 . The compliance maturity (see Equation 6.6) of a firing sequence σ_2 shows the degree to which the specification of a reference model expressed with σ_1 is followed by σ_2 .

Example 13 (Firing Sequence Compliance)

Take the following extended firing sequence set: $\sigma_i^{ext} = \{ \langle \text{Categorize incident, Prioritize incident, Search for a solution, Inform customer} \rangle \langle \text{Prioritize incident, Categorize incident, Search for a solution, Inform customer} \rangle \}$. Consider also the firing sequence $\sigma_j^{map} = \{ \langle \text{Categorize incident, Prioritize incident, Inform customer} \rangle \}$. The firing sequence compliance is $fsc(\sigma_j, \sigma_i, \mathcal{G}) = 3$. \diamond

Example 14 (Firing Sequence Compliance Degree and Maturity)

Consider the extended firing sequence sets σ_i^{ext} and σ_j^{map} from the previous example. The firing sequence compliance degree $fscd(\sigma_j, \sigma_i, \mathcal{P}, \mathcal{G})$ is $\frac{3}{3}$. It is logical that the instance σ_j^{map} of an as-is or to-be process model precisely follows the order prescribed by its reference model. The firing sequence compliance maturity $fscm(\sigma_j, \sigma_i, \mathcal{P}, \mathcal{G})$ is $\frac{3}{4}$. This means that only 75 % of instance σ_i prescribed by the reference model is being followed by instance σ_j^{map} , an instance of an as-is or to-be process model. \diamond

The compliance of individual instances, or firing sequences, has been analyzed. To meet the requirements specified in Section 6.2, all firing sequences of model M_1 and M_2 need to be determined. In Equation 6.7 and Equation 6.8, the degree and maturity of compliance express the ratio of instances that can be produced by one model *and* by the other model. From the viewpoint of compliance degree, the process model is related to the reference model; from the viewpoint of compliance maturity the reference model is related to the process model. These compliance measures return a value in interval $[0, 1]$. For example, if the cd is 1, the compliance must be at its apex since all firing sequences of model M_2 can also be produced by model M_1 .

6.4.3. Feasibility Study

Algorithm *Sequence-Based Compliance Analysis*, based on the generation of sets of firing sequences, describes the behavior of a process model. Unfortunately, the size of these sets can grow exponentially with the size of the WF-net in terms of activities. Despite its exponential growth, this section shows the applicability of algorithm *Sequence-Based Compliance Analysis*. As in Dijkman (2008), a sample of extended event-driven process

6. Measuring Process Compliance with Reference Models

chains of the SAP reference model has been used to determine whether the algorithm can be applied in practice – that is, whether or not it is feasible in terms of computation times. The SAP reference model is well described (see Curran et al. (1997) and Keller and Teufel (1998b)) and is referred to in many research papers, including Mendling et al. (2005), Rosemann and van der Aalst (2007), and Dijkman (2008). The SAP reference model collection contains over 600 process models expressed as extended event-driven process chains. It reflects version 4.6 of SAP R/3 marketed in 2000 (Mendling et al., 2008). Since it is among the most comprehensive reference models, covering over 600 business processes, these models are evidently a representative sample. The study has been performed by applying algorithm *Sequence-Based Compliance Analysis* to a subset of 126 pairs of eEPCs from the SAP reference model, these pairs having converted to workflow nets. They are put together based on similarities computed by the ProM plug-in *EPC Similarity Calculator*. These pairs are characterized with a similarity greater than 50 %. Figure 6.3 shows the percentage of model pairs for which the compliance can be computed within a given number of milliseconds on a PC. The figure shows that for all model pairs, measuring the compliance is a matter of milliseconds. Ninety percent of the process models that have been analyzed with the compliance algorithm took less than 62 milliseconds. In the experiment, the runtime of the algorithm took on average 50.50 milliseconds with a standard deviation of 9.30 milliseconds. Figure 6.4 shows the runtime per activities in the processes of a model pair. The average number of activities in these processes is 16. The weak correlation between runtime and the number of activities of a process led to the conclusion that the amount of activities found in the SAP reference models can be dealt with by algorithm *Sequence-Based Compliance Analysis*. These results show that one is confronted with exponential runtime when the complexity is measured in terms of the input size only, that is, activities. The case where exponential complexity does exist is seemingly an immense problem – however, in practice there are natural boundaries, that is, the number of activities per process model between a lower bound and an upper bound, meaning that the algorithm can be used in practice even then. Indeed, Verbeek (2004) has argued that state spaces generating a reachability graph are often feasible for systems up to 100 transitions.

An alternative for addressing complexity with regard to the input size of the algorithm

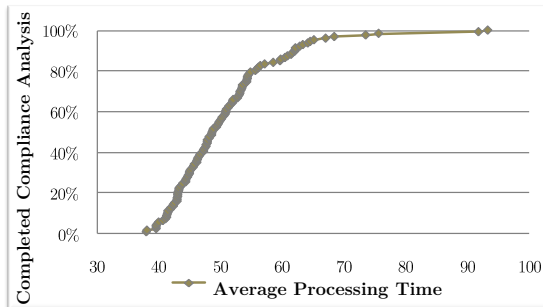


Figure 6.3.: Average runtime

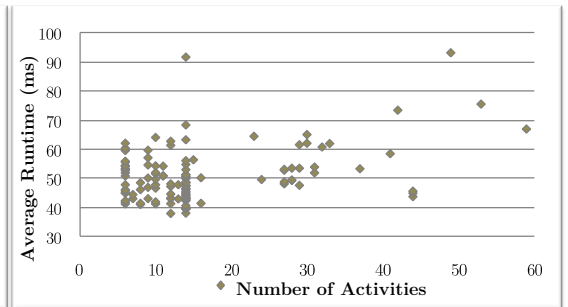


Figure 6.4.: Runtime as function of activities

is to capture the behavior of a model using the state space of a WF-net. Such a state space corresponds to the set of reachable markings of a WF-net (Basten and van der Aalst, 2001). The resulting graph is the reachability graph. Buchholz and Kemper (2002) presented a method that focused on optimizing the generation of the reachability graph of large Petri nets – one decomposes a net to generate reachability graphs for the parts and to combine them. Furthermore, there are various techniques for state space reduction (Dijkman, 2008), known as reduction rules, that may be exploited to improve the efficiency of the underlying algorithm *Sequence-Based Compliance Analysis*. These rules endeavor to reduce the size of the state space by cutting the number of places and transitions preserving information relevant for analysis purposes. For example, it is possible to account for the significance of transitions – rarely executed ones are left out when one uses abstraction or encapsulation.

6.4.4. Running Example

Model M_1 and model M_2 in Figure 6.5 depict the concept. The process models differ in three ways. First, the additional activity Z is present in model M_1 . The second difference lies in the reverse order in which activity D and activity E are executed in model M_1 and model M_2 . Third, in model M_2 , activity D is performed in parallel to the

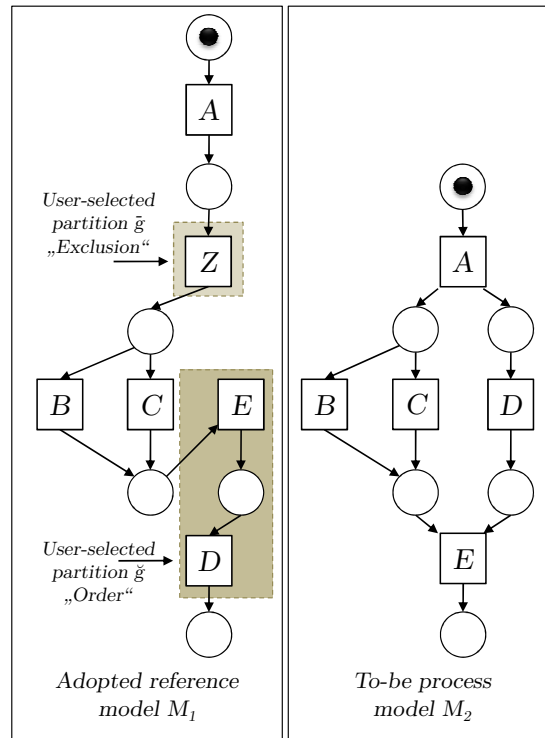


Figure 6.5.: Two process models used in the running example

6. Measuring Process Compliance with Reference Models

enactment of activity B or activity C, whereas activity D succeeds activity B or activity C in M_1 . And the models have similarities. For instance, in both models, activity B belongs to the successor of activity A and is followed by activity D and activity E. The differences and similarities indicate that the models, though not perfectly equivalent, may be compliant with each other. If they are compliant to each other, the addition of activity Z and reversing the order of activity D and activity E might not affect the compliance. With this in mind, consider the following steps:

1. Define the user-selected partitions of type exclusion (\bar{g}) and order (\check{g}) in model M_1 .
2. Determine the set of all firing sequences (S_1 and S_2) generated by model M_1 and model M_2 .
3. Derive the extended firing sequences (S_1^{ext}) of model M_1 that stem from user-selected partitions and derive the mapped firing sequences (S_2^{map}) of model M_2 .
4. Compute the fsc defined in Equation 6.4 as well as the degree of the $fscd$ and the $fscm$ according to Equations 6.5 and 6.6.
5. Select the highest $fscd$ and $fscm$ and calculate the cd and cm according to Equations 6.7 and 6.8.

The five steps seriatum:

Step 1: User-Selected Partition. A partition of type exclusion is defined as one containing activity Z ($\bar{g} = \{ Z \}$) and containing a partition of a type order embracing activity D and activity E ($\check{g} = \{ D, E \}$). In partition \check{g} , an arbitrary order of the execution of the activities is possible.

Step 2: Firing Sequences. Model M_1 generates two firing sequences $S_1 = \{ \langle A, Z, B, E, D \rangle, \langle A, Z, C, E, D \rangle \}$, while model M_2 produces four firing sequences $S_2 = \{ \langle A, C, D, E \rangle, \langle A, D, C, E \rangle, \langle A, D, B, E \rangle, \langle A, B, D, E \rangle \}$.

Step 3: Extended and Mapped Firing Sequences. As to the defined partitions, the extended firing sequences for model M_1 are $S_1^{ext} = \{ \langle A, B, E, D \rangle, \langle A, B, D, E \rangle, \langle A, C, E, D \rangle, \langle A, C, D, E \rangle \}$. Note that since the labels and functionalities of model M_1 and model M_2 resemble, a resemblance that result in $S_2 = S_2^{map}$, the mapping is omitted here.

Step 4: Fsc, Fscd, and Fscm. The fsc , $fscd$, and $fscm$ are each determined for each firing sequence in S_2 , that is, σ_{2-1} , σ_{2-2} , σ_{2-3} , and σ_{2-4} , w.r.t the four extended firing sequences σ'_{1-1} , σ''_{1-1} , σ'_{1-2} , and σ''_{1-2} (σ'_{1-1} , σ''_{1-1} , σ'_{1-2} , $\sigma''_{1-2} \in \sigma_1^{ext}(\mathcal{P})$):

$$\begin{aligned}
 fsc(\sigma_{2-1}, \sigma'_{1-1}) &= fsc(\langle A, C, D, E \rangle, \langle A, B, E, D \rangle) &= 2.00 \\
 fscd(\sigma_{2-1}, \sigma'_{1-1}) &= fscd(\langle A, C, D, E \rangle, \langle A, B, E, D \rangle) &= .50 \\
 fscm(\sigma_{2-1}, \sigma'_{1-1}) &= fscm(\langle A, C, D, E \rangle, \langle A, B, E, D \rangle) &= .50
 \end{aligned}$$

6.4. Development of a Sequence-Based Compliance Analysis

$$\begin{aligned} fsc(\sigma_{2-2}, \sigma'_{1-1}) &= fsc(\langle A, D, C, E \rangle, \langle A, B, E, D \rangle) &= 2.00 \\ fscd(\sigma_{2-2}, \sigma'_{1-1}) &= fscd(\langle A, D, C, E \rangle, \langle A, B, E, D \rangle) &= .50 \\ fscm(\sigma_{2-2}, \sigma'_{1-1}) &= fscm(\langle A, D, C, E \rangle, \langle A, B, E, D \rangle) &= .50 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-3}, \sigma'_{1-1}) &= fsc(\langle A, D, B, E \rangle, \langle A, B, E, D \rangle) &= 3.00 \\ fscd(\sigma_{2-3}, \sigma'_{1-1}) &= fscd(\langle A, D, B, E \rangle, \langle A, B, E, D \rangle) &= .75 \\ fscm(\sigma_{2-3}, \sigma'_{1-1}) &= fscm(\langle A, D, B, E \rangle, \langle A, B, E, D \rangle) &= .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-4}, \sigma'_{1-1}) &= fsc(\langle A, B, D, E \rangle, \langle A, B, E, D \rangle) &= 3.00 \\ fscd(\sigma_{2-4}, \sigma'_{1-1}) &= fscd(\langle A, B, D, E \rangle, \langle A, B, E, D \rangle) &= .75 \\ fscm(\sigma_{2-4}, \sigma'_{1-1}) &= fscm(\langle A, B, D, E \rangle, \langle A, B, E, D \rangle) &= .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-1}, \sigma''_{1-1}) &= fsc(\langle A, C, D, E \rangle, \langle A, B, D, E \rangle) &= 3.00 \\ fscd(\sigma_{2-1}, \sigma''_{1-1}) &= fscd(\langle A, C, D, E \rangle, \langle A, B, D, E \rangle) &= .75 \\ fscm(\sigma_{2-1}, \sigma''_{1-1}) &= fscm(\langle A, C, D, E \rangle, \langle A, B, D, E \rangle) &= .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-2}, \sigma''_{1-1}) &= fsc(\langle A, D, C, E \rangle, \langle A, B, D, E \rangle) &= 3.00 \\ fscd(\sigma_{2-2}, \sigma''_{1-1}) &= fscd(\langle A, D, C, E \rangle, \langle A, B, D, E \rangle) &= .75 \\ fscm(\sigma_{2-2}, \sigma''_{1-1}) &= fscm(\langle A, D, C, E \rangle, \langle A, B, D, E \rangle) &= .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-3}, \sigma''_{1-1}) &= fsc(\langle A, D, B, E \rangle, \langle A, B, D, E \rangle) &= 3.00 \\ fscd(\sigma_{2-3}, \sigma''_{1-1}) &= fscd(\langle A, D, B, E \rangle, \langle A, B, D, E \rangle) &= .75 \\ fscm(\sigma_{2-3}, \sigma''_{1-1}) &= fscm(\langle A, D, B, E \rangle, \langle A, B, D, E \rangle) &= .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-4}, \sigma''_{1-1}) &= fsc(\langle A, B, D, E \rangle, \langle A, B, D, E \rangle) &= 4.00 \\ fscd(\sigma_{2-4}, \sigma''_{1-1}) &= fscd(\langle A, B, D, E \rangle, \langle A, B, D, E \rangle) &= 1.00 \\ fscm(\sigma_{2-4}, \sigma''_{1-1}) &= fscm(\langle A, B, D, E \rangle, \langle A, B, D, E \rangle) &= 1.00 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-1}, \sigma'_{1-2}) &= fsc(\langle A, C, D, E \rangle, \langle A, C, E, D \rangle) &= 3.00 \\ fscd(\sigma_{2-1}, \sigma'_{1-2}) &= fscd(\langle A, C, D, E \rangle, \langle A, C, E, D \rangle) &= .75 \\ fscm(\sigma_{2-1}, \sigma'_{1-2}) &= fscm(\langle A, C, D, E \rangle, \langle A, C, E, D \rangle) &= .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-2}, \sigma'_{1-2}) &= fsc(\langle A, D, C, E \rangle, \langle A, C, E, D \rangle) &= 3.00 \\ fscd(\sigma_{2-2}, \sigma'_{1-2}) &= fscd(\langle A, D, C, E \rangle, \langle A, C, E, D \rangle) &= .75 \\ fscm(\sigma_{2-2}, \sigma'_{1-2}) &= fscm(\langle A, D, C, E \rangle, \langle A, C, E, D \rangle) &= .75 \end{aligned}$$

$$fsc(\sigma_{2-3}, \sigma'_{1-2}) = fsc(\langle A, D, B, E \rangle, \langle A, C, E, D \rangle) = 2.00$$

6. Measuring Process Compliance with Reference Models

$$\begin{aligned} fscd(\sigma_{2-3}, \sigma'_{1-2}) &= fscd(\langle A, D, B, E \rangle, \langle A, C, E, D \rangle) = .50 \\ fscm(\sigma_{2-3}, \sigma'_{1-2}) &= fscm(\langle A, D, B, E \rangle, \langle A, C, E, D \rangle) = .50 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-4}, \sigma'_{1-2}) &= fsc(\langle A, B, D, E \rangle, \langle A, C, E, D \rangle) = 2.00 \\ fscd(\sigma_{2-4}, \sigma'_{1-2}) &= fscd(\langle A, B, D, E \rangle, \langle A, C, E, D \rangle) = .50 \\ fscm(\sigma_{2-4}, \sigma'_{1-2}) &= fscm(\langle A, B, D, E \rangle, \langle A, C, E, D \rangle) = .50 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-1}, \sigma''_{1-2}) &= fsc(\langle A, C, D, E \rangle, \langle A, C, D, E \rangle) = 4.00 \\ fscd(\sigma_{2-1}, \sigma''_{1-2}) &= fscd(\langle A, C, D, E \rangle, \langle A, C, D, E \rangle) = 1.00 \\ fscm(\sigma_{2-1}, \sigma''_{1-2}) &= fscm(\langle A, C, D, E \rangle, \langle A, C, D, E \rangle) = 1.00 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-2}, \sigma''_{1-2}) &= fsc(\langle A, D, C, E \rangle, \langle A, C, D, E \rangle) = 3.00 \\ fscd(\sigma_{2-2}, \sigma''_{1-2}) &= fscd(\langle A, D, C, E \rangle, \langle A, C, D, E \rangle) = .75 \\ fscm(\sigma_{2-2}, \sigma''_{1-2}) &= fscm(\langle A, D, C, E \rangle, \langle A, C, D, E \rangle) = .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-3}, \sigma''_{1-2}) &= fsc(\langle A, D, B, E \rangle, \langle A, C, D, E \rangle) = 3.00 \\ fscd(\sigma_{2-3}, \sigma''_{1-2}) &= fscd(\langle A, D, B, E \rangle, \langle A, C, D, E \rangle) = .75 \\ fscm(\sigma_{2-3}, \sigma''_{1-2}) &= fscm(\langle A, D, B, E \rangle, \langle A, C, D, E \rangle) = .75 \end{aligned}$$

$$\begin{aligned} fsc(\sigma_{2-4}, \sigma''_{1-2}) &= fsc(\langle A, B, D, E \rangle, \langle A, C, D, E \rangle) = 3.00 \\ fscd(\sigma_{2-4}, \sigma''_{1-2}) &= fscd(\langle A, B, D, E \rangle, \langle A, C, D, E \rangle) = .75 \\ fscm(\sigma_{2-4}, \sigma''_{1-2}) &= fscm(\langle A, B, D, E \rangle, \langle A, C, D, E \rangle) = .75 \end{aligned}$$

Step 5: Cd and Cm. Only two firing sequences fully comply with model M_1 . Two other firing sequences partially comply with model M_1 , as can be seen from the best matching M_1 sequences for given M_2 sequence:

$$\begin{aligned} fscd(\sigma_{2-1}, \sigma''_{1-2}) &= fscd(\langle A, C, D, E \rangle, \langle A, C, D, E \rangle) = 1.00 \\ fscd(\sigma_{2-4}, \sigma''_{1-1}) &= fscd(\langle A, B, D, E \rangle, \langle A, B, D, E \rangle) = 1.00 \\ fscd(\sigma_{2-2}, \sigma'_{1-2}) &= fscd(\langle A, D, C, E \rangle, \langle A, C, E, D \rangle) = .75 \\ fscd(\sigma_{2-3}, \sigma'_{1-1}) &= fscd(\langle A, D, B, E \rangle, \langle A, B, E, D \rangle) = .75 \end{aligned}$$

From these values the compliance degree yields

$$cd(M_2, M_1, \mathcal{P}, \mathcal{G}) = \frac{3.50}{4} = .88$$

Two of the firing sequences are fully mature with given M_1 sequences, that is, $fscm(\sigma_{2-4}, \sigma''_{1-1}) = fscm(\langle A, B, D, E \rangle, \langle A, B, D, E \rangle) = 1.00$ and $fscm(\sigma_{2-1}, \sigma''_{1-2}) = (\langle A, C, E, D \rangle, \langle A, C, D, E \rangle) = 1.00$. Thus, the maximum value is given by

$$cm(M_2, M_1, \mathcal{P}, \mathcal{G}) = \frac{2.00}{2} = 1.00$$

6.5. Measuring Compliance in Practice

This section employs the *Sequence-Based Compliance Analysis* to the operational case studies introduced in Section 4.1. Each employment is accompanied by a comparison of the result with two approaches available in ProM: *Structural Precision and Recall* and *Footprint Similarity*. They have been chosen because they are commonly used to compare models, albeit rarely used in connection with compliance.

6.5.1. Complaint Handling

Figure 6.6 shows two workflow nets as starting points for the compliance analysis in ProM. The model's left-hand side portrays model M_1 , which was adopted from ITIL. Initially created as an eEPC in the ARIS toolset, it was converted to a WF-net and imported to ProM. In the figure, model M_1 is represented by a Heuristics net. The model's right-hand side displays model M_3 that represents the complaint handling process of an airline (i.e., use case Air-2). The extraction of the as-is model has been described in Section 4.2.3.

To adapt the reference model to the airline's need, model M_1 was customized as follows. The activity *Identify responsible employee* was excluded because it was not recorded by the IS. The airline presumed that the activities *Inform customer* and *Preprocess incident* could be executed in either order. As a result, the airline chose a user-selected partition of type exclusion $\bar{p} = \{ \textit{Identify responsible employee} \}$ and a user-selected partition of type order $\check{p} = \{ \textit{Inform customer}, \textit{Preprocess incident} \}$. The left-hand side of Figure 6.7 shows the granularity mapping. It is important to note that the figure depicts the as-is model as model M_2 though the text refers to model M_3 ⁸¹. During the mapping, the airline process, discussed in Section 6.2.2 manifested typically these characteristics: missing and additional activities as well as activities with different levels of detail. For example, the activity *Prioritize incident* is missing in model M_3 and the activities *Create activity customer relations* and *Create activity customer payments* of model M_3 correspond to the activity *Create incident* in model M_1 . Figure 6.6 shows that the airline uses iterations – model M_3 has cycles. Since the cycles had been viewed as indications of quality improvement, the limit for cycle unrolling was set to 1, ensuring both that all activities were considered and that the activities' iteration has had no adverse effect.

Figure 6.7's right-hand side illustrates the results of the compliance analysis. Both the compliance degree and compliance maturity were computed with Equations 6.7 and 6.8 per passed cycle as well as the extended firing sequences $\sigma_1^{\text{ext}}(\mathcal{P})$ of model M_1 and the firing sequences $\sigma_3^{\text{map}}(\mathcal{G})$ of model M_3 . Running through a cycle once yielded the compliance degree $cd(M_3, M_1, \mathcal{P}, \mathcal{G})$ of .82 and the compliance maturity $cm(M_3, M_1, \mathcal{P}, \mathcal{G})$ of .52. The first line of the sequences σ_1 and σ_3 explains these values. Consider the following extended firing sequence $\sigma'_{1-1} = \langle \textit{Receive incident}, \textit{Identify account}, \textit{Create incident record}, \textit{Process incident}, \textit{Categorize incident}, \textit{Prioritize incident}, \textit{Search for}$

⁸¹ The algorithm, by definition, compares a reference model M_1 with a process model M_2 that can be the as-is model herein referred to as model M_3 .

6. Measuring Process Compliance with Reference Models

a solution, Make solution available, Inform customer, Preprocess incident, Close incident) and $\sigma''_{1-1} = \langle \text{Receive incident, Identify account, Create incident record, Process incident, Categorize incident, Prioritize incident, Search for a solution, Make solution available, Preprocess incident, Inform customer, Close incident} \rangle$, $\sigma'_{1-1}, \sigma''_{1-1} \in \sigma_1^{\text{ext}}(\mathcal{P})$. Consider the firing sequence $\sigma_{3-1} = \langle \text{Open complaint, Receive contact, Edit mail, Classify problem, Identify account, Create activity Cust. Relations, System allocates flight data, Close complaint} \rangle$, which resulted in the firing sequence $\sigma'_{3-1} = \langle \text{Receive incident, Categorize incident, Identify account, Create incident record, Process incident, Close incident} \rangle \in \sigma_3^{\text{map}}(\mathcal{G})$. Since the maximum lcs of σ'_{1-1} and σ''_{1-1} with σ'_{3-1} corresponded to $\langle \text{Receive incident, Identify account, Create incident record, Process incident, Close incident} \rangle$, the firing sequence compliance $fsc(\sigma_{3-1}, \sigma_{1-1}, \mathcal{P}, \mathcal{G})$ was five. The firing sequence compliance degree $fscd(\sigma_{3-1}, \sigma_{1-1}, \mathcal{P}, \mathcal{G})$ was $\frac{5}{6}$. It follows that the instance σ_{3-1}

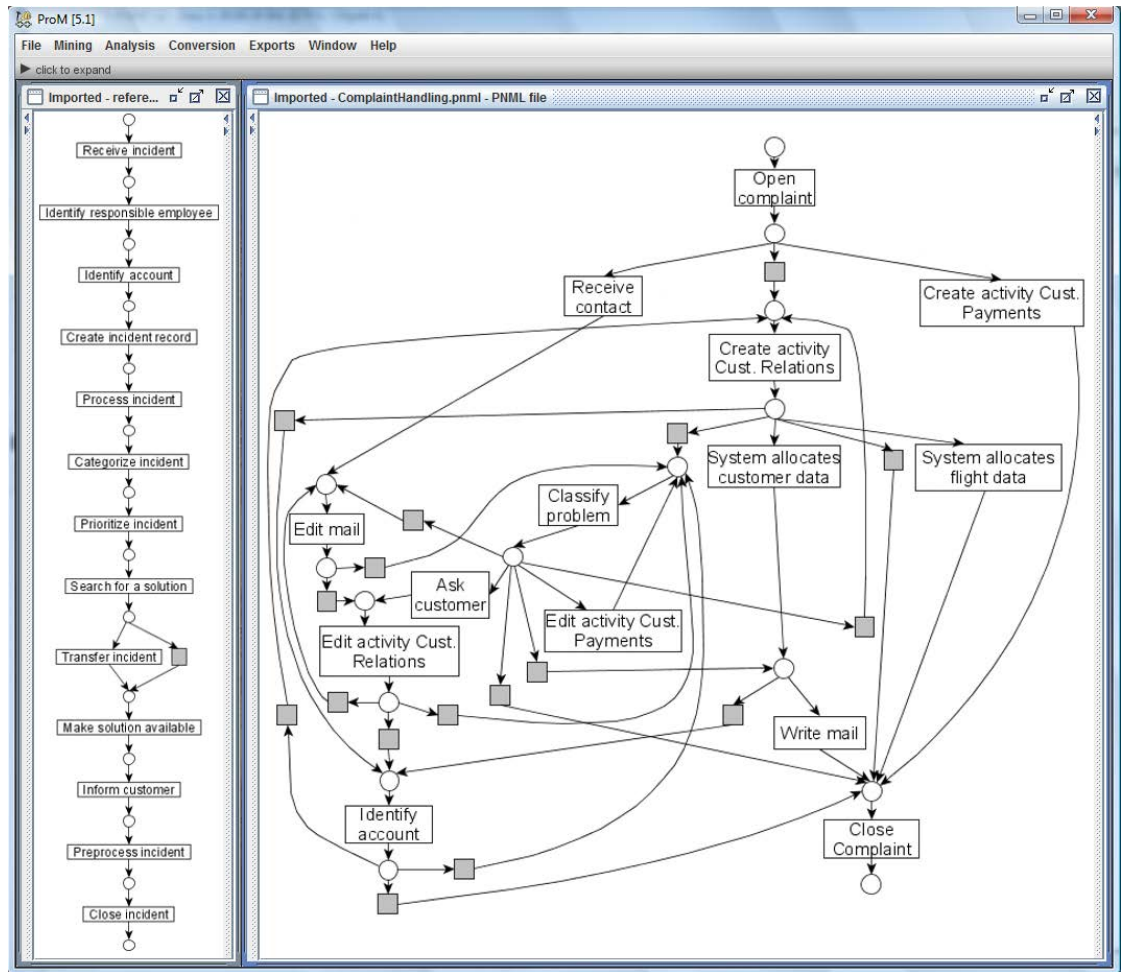


Figure 6.6.: Process model Air-2 used for compliance analysis

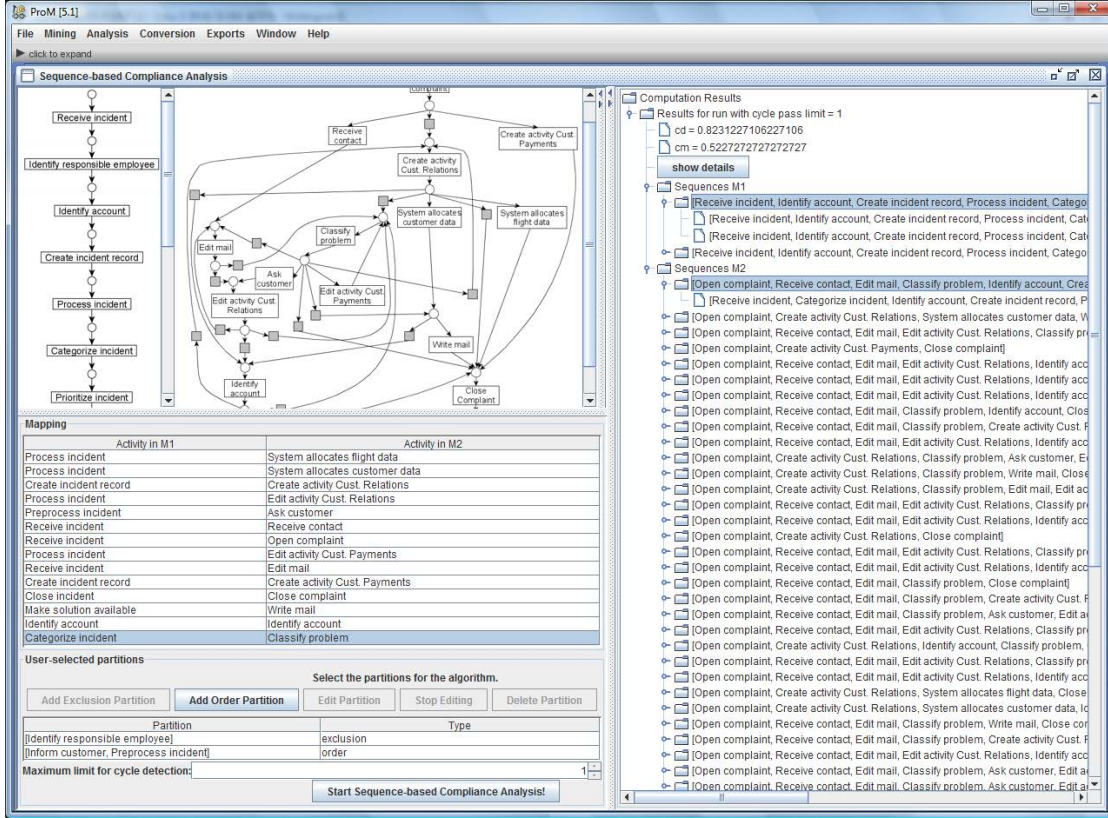


Figure 6.7.: Sequence-Based Compliance Analysis plug-in

of the as-is process model preserved the order of the reference model with an overlap of 83 %. The firing sequence compliance maturity $fscm(\sigma_{3-1}, \sigma_{1-1}, \mathcal{P}, \mathcal{G})$ was $\frac{5}{11}$. This means that only approximately 45 % of instance σ_{1-1} prescribed by the reference model was followed by instance σ_{3-1} of the as-is process model.

The compliance degree of .82 displays a high correspondence of the processes executed by the airline with the recommendations of the reference model. Although models M_1 and M_3 appear to be substantially different, the model M_3 is highly compliant with model M_1 . The compliance maturity of .52 means that there are recommendations in model M_1 that have not been implemented by the airline. The maturity value of .52 indicates that model M_3 is also partially mature with model M_1 .

A comparison of case Air-2's *Sequence-Based Compliance Analysis* plug-in with both the *Structural Precision and Recall* plug-in and the *Footprint Similarity* plug-in follows.

As to *Structural Precision and Recall*, models M_1 and M_3 must be represented by a Heuristics net. Therefore, model M_1 , originally represented by an eEPC, was converted to a Heuristics net using ProM. Since the ProM plug-in expects same labels for activities representing the same functionalities, the labels of model M_3 were renamed according to

6. Measuring Process Compliance with Reference Models

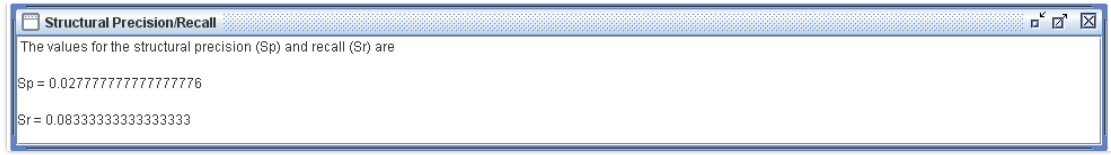


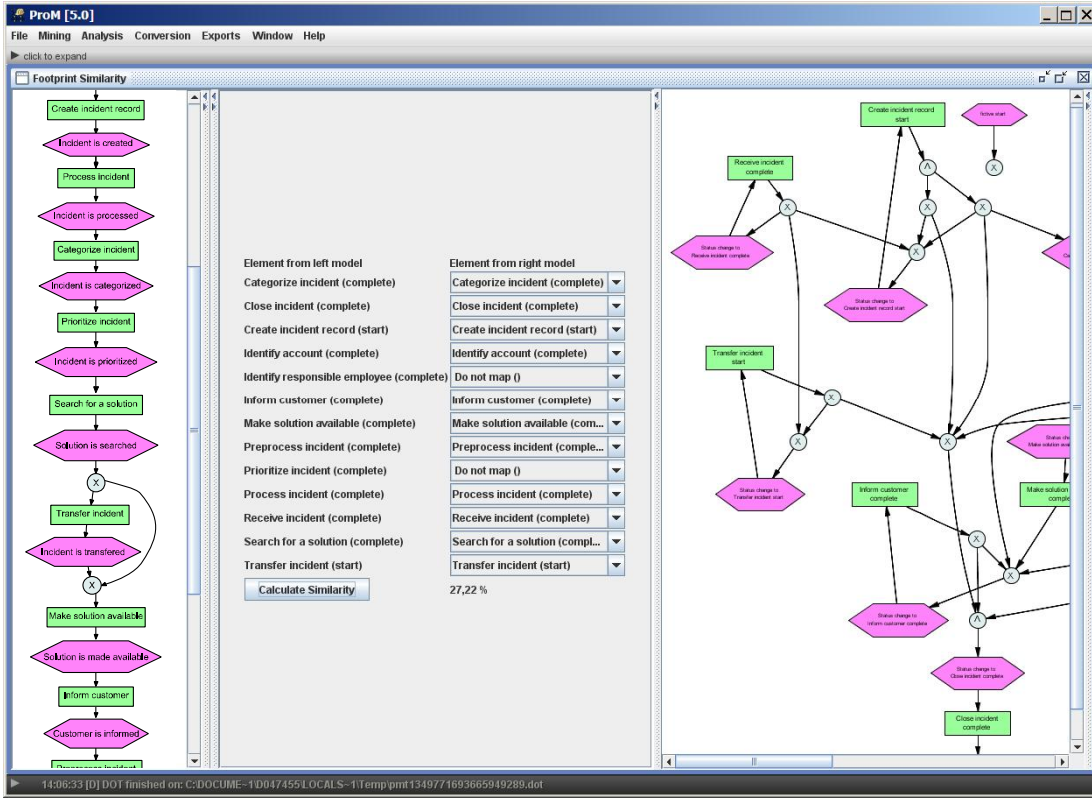
Figure 6.8.: *Structural Precision and Recall* plug-in Air-1

model M_1 and the mapping was carried out as depicted in Figure 6.7. Figure 6.8 depicts the results of the comparison between models M_1 and M_3 . The *Structural Precision* was .03 and the *Structural Recall* was .08.

The *Footprint Similarity* is the second approach the algorithm is compared with. Since the analysis of the *Footprint Similarity* is based on comparing of two eEPCs, a conversion plug-in in ProM converted model M_3 to an eEPC. The mapping was done manually – see Figure 6.7. To analyze the *Footprint Similarity*, the ProM plug-in *Footprint Similarity* was used, yielding a result of .27 (see Figure 6.9).

The compliance values warrant detailed discussion. *Structural Precision and Recall* rely on the notion of equivalence – they expect process models to be equal in their structure, an expectation consonant with the goal of unifying multiple views of a process (e.g., harmonization after an organizational merger), though this expectation is dissonant with compliance. The processes resemble each other though their structures differ. Therefore the values obtained .03 and .08 were relatively low. The new approach like the previous two, provides guidance that facilitates the analysis of compliance from the perspectives compliance degree (i.e., *Structural Precision*) and compliance maturity (i.e., *Structural Recall*). But the previous two neither offer a mapping functionality nor address the necessary customization (ordering or exclusion of activities) of the reference model. Expressing the behavior of a model in terms of connections both results in information loss whether or not two connected transitions are part of a cycle, and neglects the control flow of process models – these two detriments result in the absence of much-needed information when measuring the compliance with reference models.

The *Footprint Similarity* also relies on the notion of equivalence, while not insisting that two models are identical – this approach assumes that process models with different structures nevertheless may be similar. Therefore the result of .27 is closer to the values obtained when using the algorithm herein developed (i.e., .82 and .52). Since the formula is symmetric, measuring the compliance of model M_3 with model M_1 or of model M_1 with model M_3 yields the same value. While patently this situation is absolutely consonant with the notion of equivalence, it fails to meet the requirements that compliance ought to be determined as to degree and maturity. As in the case of algorithm *Sequence-Based Compliance Analysis*, the notion of mapping is here included as well. However, a non-injective mapping is not supported. Since the algorithm *Footprint Similarity* addresses the ordering of activities, it partially fulfills the requirement for customization of reference models – it does not address the exclusion of activities. Van Dongen et al. (2008) did not discuss the ramifications of employing their formula with respect to cycles.

Figure 6.9.: *Footprint Similarity* plug-in of Air-1

The employment of algorithms predicated on the notion of equivalence tempts one to infer that processes are not compliant. In contrast to the *Sequence-Based Compliance Analysis*, *Structural Precision and Recall*, and *Footprint Similarity* each gives a value that is difficult to explain, and one with which one struggles to commit oneself to a characterization. It is possible to trace neither the missing nor the non-compliant instances. The solution proposed in this dissertation provides two different values for compliance, namely, degree and maturity, and calculates intermediate results from instance compliance. These enable process designers to trace the instances that positively or negatively affect the compliance of the processes being analyzed. Applications in actual business situations have proven that the notion of equivalence cannot be used with satisfactory results to evaluate the compliance of processes with a reference model.

6.5.2. Service Operation

Figure 6.10 shows the *Sequence-Based Compliance Analysis* plug-in of use case Tel-1, which was inputted with the reference model M_1 (Figure 5.17) and the as-is model M'_3 (see Figure 4.25's right-hand side model). Model M_1 generated one sequence, while

6. Measuring Process Compliance with Reference Models

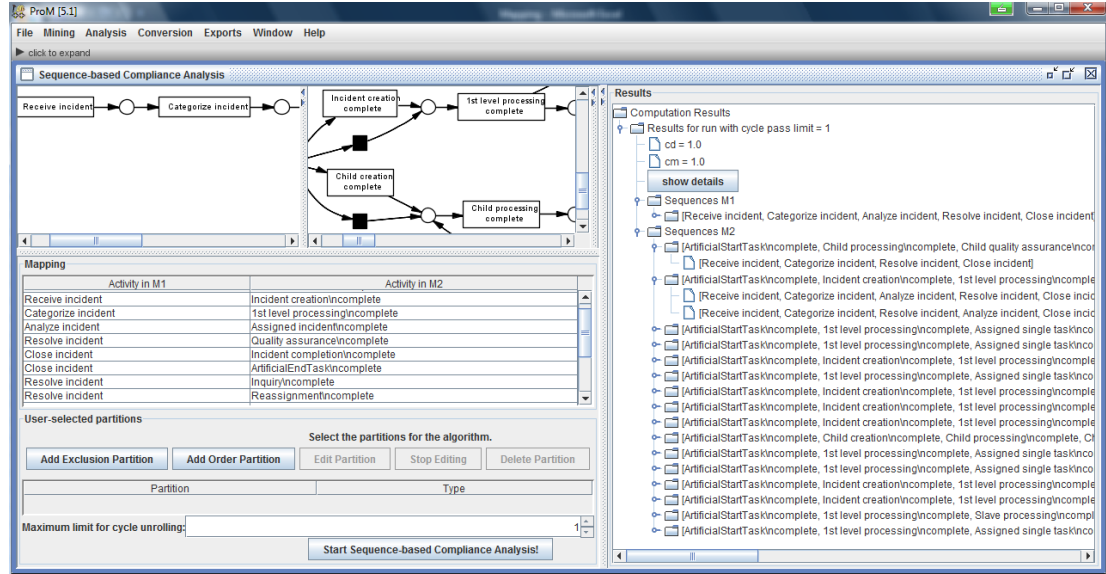


Figure 6.10.: *Sequence-Based Compliance Analysis* of Tel-1

model M'_3 generated 16 sequences. Because of the values of 1.00 and, again, 1.00 as to compliance degree and compliance maturity respectively, it is certain that ISP's processes followed the recommendations of the underlying model M_1 and certain, too that the specifications of M_1 were adopted absolutely.

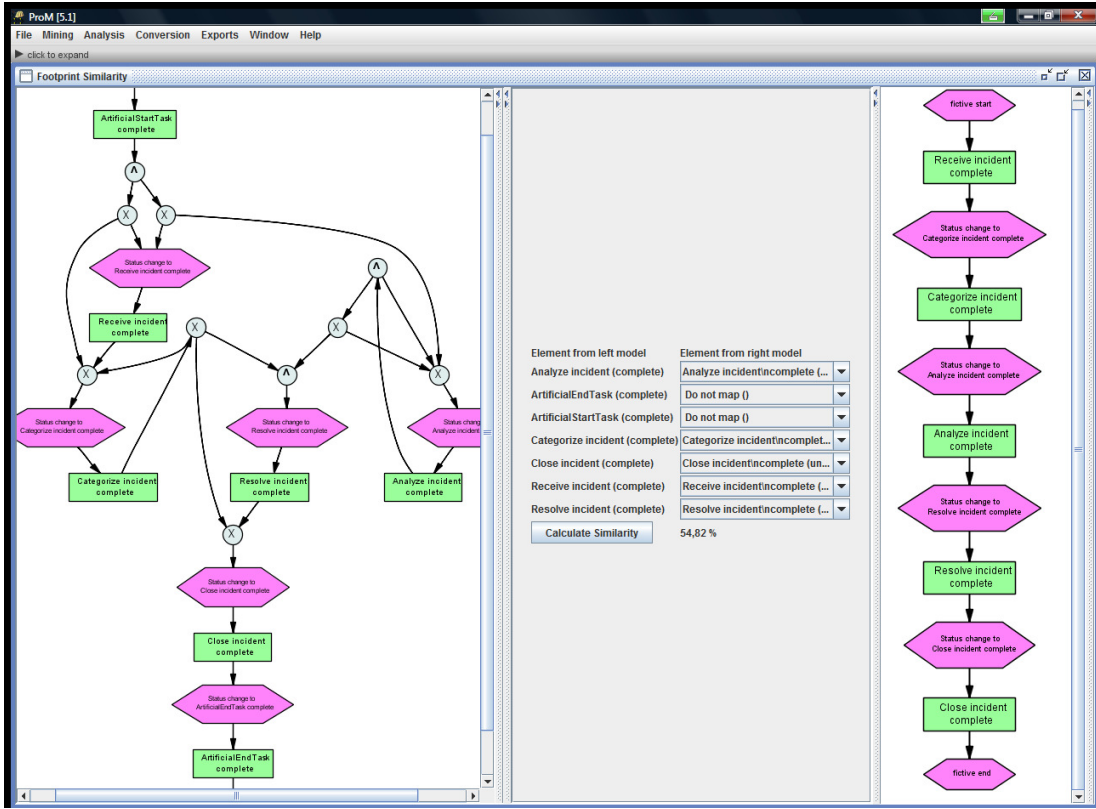
In contrast, the *Structural Precision* yielded a value of .47 and the *Structural Recall* that of .75. According to the *Footprint Similarity*, the to-be model and the as-is model bore a similarity of .55, as shown in Figure 6.11. (Section 6.5.1 has discussed the reasons for the deviations between the *Sequence-Based Compliance Analysis* and the two comparative values.)

6.5.3. Pharmaceutical Drug Delivery

Figure 6.12 depicts the *Sequence-Based Compliance Analysis* plug-in of use case Pha-1. There were three sequences in the to-be model on the left-hand side (i.e., M_1) and 589 sequences in the as-is model on the right-hand side (i.e., M_3). The compliance degree was .99 – the compliance maturity was .89. These values were obtained without user-defined partition being provided.

Figure 6.13 shows the respective lists comprising the best matching sequences. The upper part of the figure shows the end of the list's final entries as to the best matching M_1 sequences for given M_3 sequences, the lower part of the figure lists the best matching M_3 sequences for given M_1 sequences.

With reference to the *fscm*, one sequence of model M_3 was found that complied perfectly with model M_1 , namely the branch starting at Athlone. The other two values

Figure 6.11.: *Footprint Similarity* plug-in Tel-1

were slightly below the value of 1.00. This very small deviation of branch $\langle Tjoapack, Actavis, HCL (Avon), HCL (HW), UniChem, Barts \rangle$ from the specification $\langle Actavis, Tjoapack, Actavis, HCL (Avon), HCL (HW), UniChem, Barts \rangle$ can be explained by the fact that all items were serialized at Tjoapack. Since the observation of the events can only start after serialization, the recorded product flow was deferred to Tjoapack. One user-selected partition, defined by type *exclusion* including the activity *ACTAVIS*, raised the *fscm* from .86 to 1.00, and with that the *cm* increased as well to .93, too. For the purpose of explaining the third sequence of model M_3 , the *fscd* values in the upper lists are noteworthy. Eleven M_3 sequences were a *fscd* < 1. Lines three (*fscd* = .80) and five (*fscd* = .75) exemplify those sequences, all of which are Sandoz's sequences, particularly those, which began at Tjoapack and then went to Sandoz. There were eleven more M_3 sequences that were involved of the Sandoz branch, which directly started at Sandoz (packed at Tjoapack), that had never seen Tjoapack (all with an *fscd* = 1.00). The problem might therefore be attributed to an incorrect ordering between Tjoapack and Sandoz. Adding a partition of type *order* $\check{p} = \{ Sandoz, Tjoapack \}$ in fact increased both the *cd* and the *cm* to the maximum value of 1.00.

6. Measuring Process Compliance with Reference Models

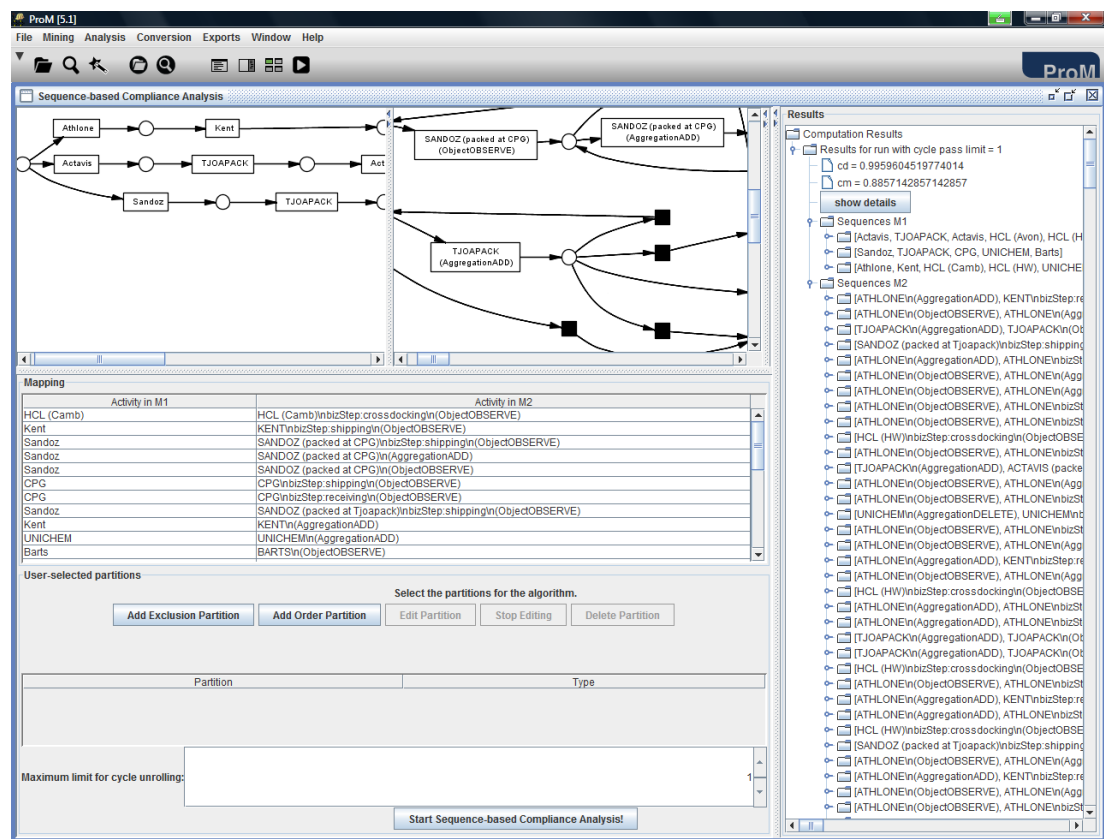


Figure 6.12.: *Sequence-Based Compliance Analysis* in Pha-1

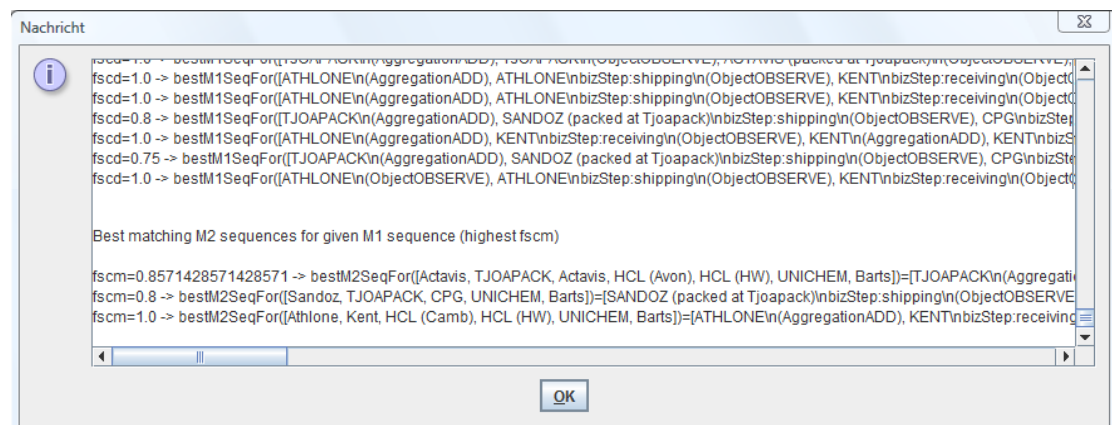
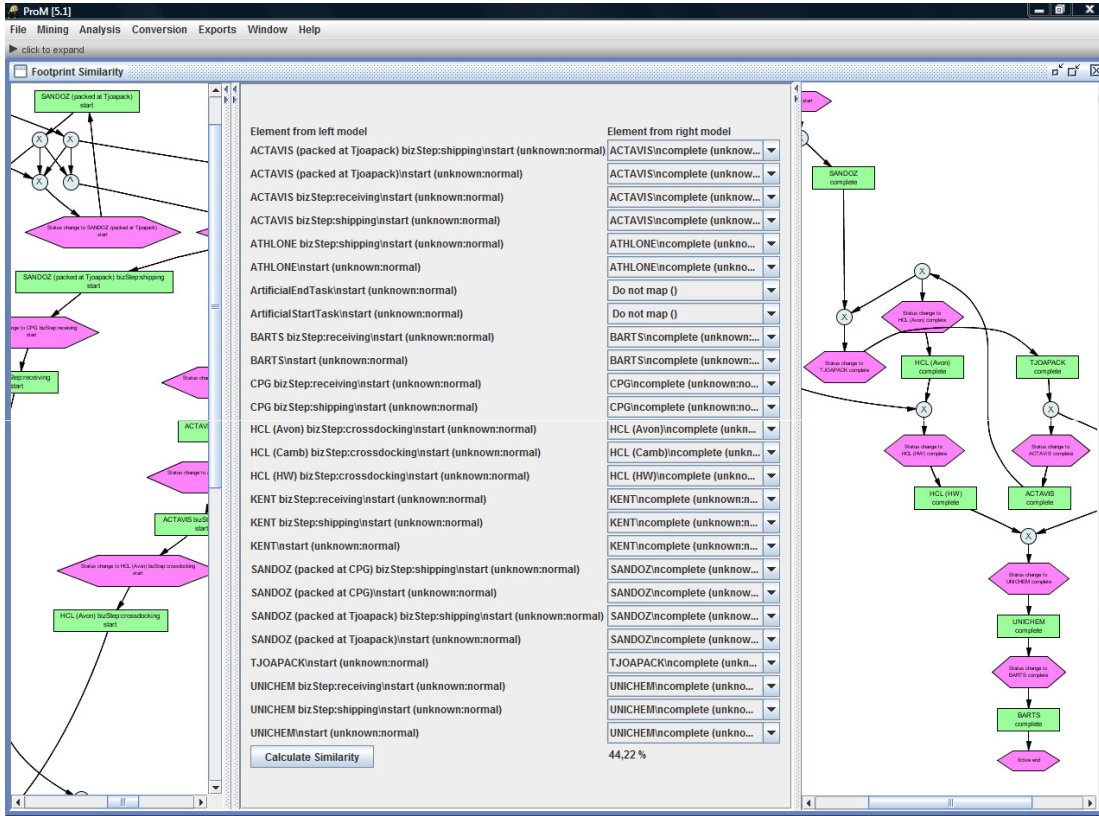


Figure 6.13.: *Sequence-Based Compliance Analysis* in Pha-1 - details

Figure 6.14.: *Footprint Similarity* plug-in Pha-1

In contrast, the *Structural Precision* reveals a similarity between the two process models of .31 and that of the *Structural Recall* of .78. Figure 6.14 depicts the process models' having a greater similarity, that of .44 this as a result of the *Footprint Similarity*. The differences between the values of similarity in the behavior of the process models are just like the ones explained in Section 6.5.1.

6.6. Conclusion

Reference models provide valuable recommendations for the implementation of business processes. Methods and solutions to determine *how* these guidelines are implemented in practice are non-existent. Extant algorithms to evaluate the equivalence of processes have proven to be inadequate for measuring compliance since many factors and characteristics related to compliance are ignored. This dissertation has thus far investigated the characteristics of compliance and has devised a generic approach to analyze the compliance of process models with reference models. The principal contribution herein is an algorithm called *Sequence-Based Compliance Analysis* that is based on the fact

6. Measuring Process Compliance with Reference Models

that process models can have different structures but one process can nevertheless be compliant with each other.

Customization of the reference model allows one to account for different granularities, user-selected partitions, mapping, and cycles that typically need to be addressed when comparing a process model with a reference model.

Two measures are offered by the *Sequence-Based Compliance Analysis*: *compliance degree* and *compliance maturity*. The *cd* measures the degree to which the processes of an organization follows the behavior recommended by the reference model. The *cm* measures the degree to which the behavior recommended by the reference model is implemented by the organization. Both measures provide valuable information, and each measure vies with the other for greater importance in a real-world business setting. Consider that organizations rarely use a reference model to its full extent – the *cm* naturally falls below the maximum. The same applies for a delta analysis. It may well be that in these two settings the *cd* value is of more importance. Note that virtually all organizations incrementally adopt a reference model. Therefore, the focus may turn to the *cm* that is expected to increase progressively with each stage in the model's adoption.

In order to validate concept and implementation, the use cases' processes have been analyzed to determine to what extent they comply with the reference model ITIL. The result has been juxtaposed with the results obtained when employing the approaches *Structural Precision and Recall* and *Footprint Similarity*. This validation has not been merely an academic exercise as process mining and equivalence algorithms have been employed on real data. The results have shown that the *Sequence-Based Compliance Analysis* has yielded values much more applicable in real business situations for measuring compliance than the results of the algorithms based on analyzing the equivalence of processes.

Future researchers may wish to investigate further the contribution of cycles to compliance with reference models. Large questions to be addressed include those about the semantics that determine how cycles contribute to compliance, about the options of which one may avail oneself in view of the semantics, as well as those about an optimal unroll cycle. Seek to learn which additional types of customization of reference models are important and to study how traceability can be incorporated into compliance analysis to enable organizations to identify swiftly problematic parts of their running processes.

7. Conclusion

This dissertation has presented a methodology for continually improving processes. A procedure model for the CPSI has been developed, one which detects deviations from reference models through continually monitoring the behavior of process execution and verifying it against specifications. Concepts and implementations, augmented the CPSI, have made it possible to employ it to real-world business settings, establishing its usefulness.

This chapter both summarizes the results and notes the contributions and lessons learned. Naturally, there are both general challenges in this area as well as specific future endeavors that may extrapolate from the work herein, and this chapter will highlight the findings' significance.

7.1. Results

The dissertation has covered a broad spectrum of research activities each of which has targeted process improvement and augmenting the CPSI.

Contributions to the theoretical framework and methodology have been principally inspired by five domain-spanning case studies conducted at a variety of organizations. The case study research method has been chosen because it offers a coherent research strategy for fathoming real-life challenges in process management.

A new methodology has been developed that actively supports the analysis, monitoring, and control process for both ITSM processes and business processes. Based on the reference model, to-be processes are set up and KPIs are determined. As-is processes and their KPIs, derived by process mining, are compared to these to-be processes. The CPSI is predicated on the use of process mining and the application of the ITIL recommended seven-step procedure. The CPSI identifies and corrects deviations from reference processes and it determines and rectifies service-specific weaknesses of the process implementation. The CPSI integrates processes, personnel, and resources.

By employing the case studies, an empirical analysis of process mining has identified two obstacles to broader use of the CPSI: only process-oriented data can be used and there are missing tools for measuring compliance of process models with reference models. Both of these have been addressed. As to that limitation, EPCIS events and enterprise data – as new data carriers – have been made accessible for process mining. The algorithms *EPCIS2MXML* and *BO2MXML* successfully grapple with the challenge of transforming RFID events and enterprise data to traceable cases, the activities of which had to be made comparable to those of a reference model vis-a-vis granularity⁸².

⁸² This is true both for new data sources and for data in workflow management systems.

7. Conclusion

As to those missing tools, the analysis of characteristics of compliance has led to a widely applicable approach that expresses the degree of process compliance and process maturity between two process models one of which can be a reference model. Algorithm *Sequence-Based Compliance Analysis*, developed herein, is based on the observation that process models can have different structures but one process can still be compliant with another.

The concepts, once thought through, have been implemented in actual business situations, and these implementations can be processed in ProM, and have been applied in all use cases to prove the concepts' rectitude.

7.2. Lessons Learned

Five insights have been garnered in the process of research with respect to process mining's appropriateness for the CPSI, to the attractiveness of RFID and enterprise data as well as to the evaluation process mining's limitation to process-oriented data, to the challenges of compliance measurement, to the algorithms' usefulness to check compliance, and to the outcomes of the CPSI.

First, process mining has proven to be a valid and powerful means to support successful reengineering efforts. Process mining for continual process improvement is most welcome because the technique is objective and can be used in a highly automated way – these two features provide reproducible and repeatable results, which are naturally essential both for giving an accurate account of the as-is condition and for keeping process models up-to-date. That said, note that it has been shown herein that process mining algorithms are, at this writing, restricted to MXML, a format that does not take into account business related information (such as documents attached to an activity), thereby reducing the expressiveness of the resulting process model. And process mining can only be transposed to case-oriented processes. Undeniable, process mining projects the behavior of an IS onto that of an organization as a result of which information outside the IS lays beyond the scope of process mining.

Second, it has been shown that the EPCglobal standard can be used to make RFID events accessible for process mining, even inter-organizational process mining. This demonstrates that process mining is employable not only to process-aware information systems and to single-systems, but can be implemented more comprehensively, notably among supply chain partners. Algorithm *EPCIS2MXML* reconstructs single EPCIS events to traceable products' movements, thereby allowing *product flow* models to be derived. It has been demonstrated that enterprise data, like RFID events, is a valuable source for process mining. The automated reconstruction of the document flow by algorithm *BO2MXML* substantially reduces the effort for manual collection. The document flow, not being restricted to one particular enterprise system, can be used in a number of applications, such as purchasing and sales processing.

Third, it is crucial to note that the comparability of activities' granularity characterizing their level of detail is a major factor for measuring compliance among activities of a process model and a reference model. As to the data's nature, transactions in enter-

prise systems have been found to be coarse-grained and events in workflow management systems have been found to be fine-grained, thereby rendering them incomparable to the reference model's activities. Process mining has shown its potential for identifying the underlying transaction scheme of the WFMS that allows for generalization from information inherent in this scheme, thereby putting the WFMS events on the same level of granularity as the reference activities. It is encouraging that it has been possible to decompose transactional enterprise data, though in a number of cases the activities' execution times have been only an approximation because they have been the ones stamped in the corresponding transaction.

Fourth, current algorithms to evaluate the equivalence of processes in the process mining context, sound in themselves, have proven to be inadequate for measuring compliance since many factors and characteristics related to compliance, such as different granularities and iterations of process activities, have been ignored. The compliance approach herein accounts for two facts, that process mining does project the behavior of an IS onto that of an organization, and that a widespread over-reliance on sequential adherence within executed activity is prevalent. Juxtaposing compliance measures of different approaches has shown that the *Sequence-Based Compliance Analysis* has yielded values much more useful in real business situations for measuring compliance than the results of the algorithms based on analyzing the equivalence of processes.

Fifth, the case studies' organizations have clearly benefited from the use of the CPSI. They have obtained transparency of both their business and ITSM processes and their processes' alignment with reference models. Most noteworthy, it has been shown that the CPSI not only ensures that business processes and ITSM processes function exactly as specified but identifies and corrects service-specific weaknesses of the process implementation. The CPSI is also concerned with the most effective use of limited resources in terms of personnel and tools. It efficiently deals with the inevitable multitude of management tasks such as process modeling, documenting, and analyzing. The CPSI is also apt because it adverts to changing business requirements and deviations both between key indicator values and their target values and between the as-is process model and the to-be process model. The approach serves as a basis for post-control of measures taken with the intention of bringing about a business goal. Process quality can be measured and controlled through quantifiable information. Measurement is reproducible, repeatable, and comparable, these three providing a base for improvement measures and post-control. The approach's high level of automation contributes to process maturity. The results have shown that the CPSI brings much aggrandizement in process quality and maturity and fills current gaps in management instruments. Concernedly, process mining results in the limitation that the CPSI can only be transposed to data prepared according to MXML and corresponding requirements.

7.3. Significance of Findings

The dissertation has as its subject the continual improvement of processes, which has been ranked by Gartner Research (2010) among the top three business priorities for

7. Conclusion

2010⁸³. This dissertation's findings, summarized below, indeed do contribute to the increased use of analytics and to the improvement of business processes.

Anticipated benefits from RFID technologies and related standards in terms of efficiencies in supply chain operations had been discussed prior to this dissertation, while strategic potentials had generally been neglected. The value of providing precise data on product movements lies increasingly in advanced *analytics* that empower greater flexibility in decision making during and after process implementation and execution. Especially within the process mining context, this information allows for thorough process analysis for tracking and tracing anomalies. One finding to emerge is that EPCIS events truly enable supply chain analysis.

Despite potential benefits vis-a-vis *process improvement*, neither reference models nor process mining have yet to be broadly adopted in industry, and they are virtually never used in tandem. This work has grappled with the impediments to their adoption. First, the provision of the two new data sources, RFID events and enterprise data, naturally has made process mining more attractive to business. The quality measures *cd* and *cm* of algorithm *Sequence-Based Compliance Analysis*, showing their relevance commensurate with the degree of compliance with specifications, provide guidance to organizations vis-a-vis reference models' adoption and implementation in a specific business context. Second, business persons have been aided in their appraisal of process mining's potential, especially with respect to its capability for improving processes by determining quality-reducing and quality-improving factors. Third, reference models and process mining have been integrated in the overall CPSI, an integration from which both BPM and ITSM may benefit.

The CPSI's validation that IT processes are in line with business requirements is an effective method to attain greater IT business alignment. The CPSI enhances business processes, patently a manifestation of IT's value, for CPSI ensures compliance with reference models assessed to be the most important instrument for IT governance. Therefore the CPSI has markedly very substantial positive effects on IT governance – a means for *improving processes*.

The CPSI is concerned with the overall life cycle of BPM. An organization availing itself of it has the realistic potential to automate tremendously process *analysis*, process modeling, process monitoring, and process evaluation. Process mining's potentials for process design and modeling challenge the traditional process modeling methods. Modeling becomes independent of the process experts' conceptions. The gap between normative modeling for compliance purposes and the actual execution of a workflow shows this. In the hitherto small number of cases in which process mining has been applied in the domain of BPM, virtually all approaches have been related to BPR. Rather than striving for radical change, one using the CPSI produces incremental improvements in attaining goals by swiftly adapting to evolving conditions, thereby increasing flexibility. That is, the CPSI takes current business or ITSM processes as a starting point and gradually refines them. Therefore the CPSI resembles the CPI because each is concerned

⁸³ The topic relates to the business priorities *improving business processes* and *increasing the use of information/analytics*, both of which will be very shortly addressed.

with quality enhancement – each *improves processes* – but they differ in that the CPSI focuses on IT supported processes.

7.4. Future Issues and Directions

The findings of this dissertation suggest three courses of action – one each for the inter-organizational use of the CPSI, for the standardization of log files, and for the stimulation of extended functionalities of the CPSI.

First, a contribution has been made to supply chain analysis, which in turn has facilitated addressing the consolidation of log files from disparate information systems and distributed supply chain parties. Future endeavors should include data consolidation and data aggregation as well as data sharing in distributed environments.

Second, recent developments in the field of process mining have led to a revival in standardization efforts for logging event data. Researchers may wish to focus their efforts on accounting for business related information, such as documents and involved supply chain parties, in order to deal with the complexity inherent in a business process. The appropriate articulation of TransactionEvents needing to be related to an activity would be especially desirable.

Third, research in extending CPSI's functionalities ought to be undertaken. Methods must be developed to build a knowledge base as an input for pattern analysis - information about prior deviations, such as solution, type, and reason, could flow in such a base and the pattern analysis could automatically classify deviations and simultaneously suggest solutions. This information could be useful to provide simulation and concomitant forecasts to examine alternative scenarios. Future challenges also lie in addressing process compliance of processes with reference models during their execution.

Appendices

A. Questionnaire

For the implementation of effective and efficient business process management, organizations need to analyze current performance. The results of the analysis show where and how improvements can be made. This applies to an holistic approach which includes business processes supported by IT and processes managed within IT.

About the Survey

This survey has the objective to analyze the view of practitioners from industry on IT management activities in the light of IT's contribution to business process improvement.

Topics Covered

The general topics covered by the questions included:

1. Importance of IT and IT management
2. Alignment of IT management and business process management
3. Effectiveness of IT management frameworks (i. e., reference models)
4. Effectiveness of process mining techniques

Time: It takes some 30 minutes to participate in this survey.

Confidentiality

This survey asks your opinion about specific aspects of IT management and related activities. It may be seen only and may be used only for statistical purposes. Your individual responses will remain confidential and will be compiled with those of other interviewees.

A. Questionnaire

A.1. Importance and Performance of IT and IT Management

Please provide us with your overall evaluation of the importance of IT and the management of related IT activities.

1. Who is responsible for the IT management in your organization? <i>(Please check one)</i>				
IT manager	<input type="checkbox"/>			
CIO	<input type="checkbox"/>			
Other <i>(Please specify)</i>	<input type="checkbox"/>			
2. How important do you, yourself, consider ...		Not very important	Very important	Don't know
...IT's contribution to business efficiency?		1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
...IT's alignment with the overall IT strategy?		4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
...IT's alignment with the overall business strategy?		1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
...IT's contribution to business innovation?		4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
...IT's management's contribution to the IT vision?		1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
		4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
3. Resources committed to process improvement/communication with business about processes within the last 12 months include: ... (Specify the number of people)				
		Improve processes	Communicate with business	Don't know
... Full-time involvement	 FTE ⁸⁴ FTE	<input type="checkbox"/>
... External consultants (in average)	 FTE FTE	<input type="checkbox"/>
Other <i>(Please specify)</i> FTE FTE	<input type="checkbox"/>

Please provide us with your overall IT strategy.

4. How would you rate your organization's maturity on IT management? By maturity, we mean the degree to which the IT is explicitly defined, managed, measured, and controlled. <i>(Please check one)</i>	
We do not think this needs attention.	<input type="checkbox"/>
We understand this is an issue but are just starting to assess what needs to be done.	<input type="checkbox"/>
We know that this is important and we have a number of performance.	<input type="checkbox"/>
We have well-defined IT management measures and processes.	<input type="checkbox"/>
We have well-functioning IT management processes and a measuring system.	<input type="checkbox"/>
Our IT management processes are continuously optimized based on measures.	<input type="checkbox"/>
Don't know	<input type="checkbox"/>

⁸⁴ Full-time equivalent

A.1. Importance and Performance of IT and IT Management

5. To what extent would you agree or disagree that the following are important to measure IT performance?	Not very important	Very important	Don't know
IT operational performance improvement	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Appropriate use of IT resources	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Degree of adoption of best practices	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Customer satisfaction (internal/external)	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
IT process maturity	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Quality of IT products/services	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
IT-related risk analysis	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Other (<i>Please specify</i>)	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>

6. Please characterize the behavior of your IT department. We provide two contrasting situations from which you						2	1	0	1	2	can choose.
Our focus on IT is ...											
... technology-oriented.				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... client-/customer-focused.		
... function-oriented.				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... process-oriented.		
... achieving competitive advantage through the use of IT.				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... guaranteeing cost-efficient production of business products.		
Our expectation about IT is that ...											
... requirements are met according to specific IT projects and processes.				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... its value is comprehensible and is subject to a continuous improvement.		
Our process of evaluating IT performance is based on ... We distinguish four main performance aspects, which can be targeted for improvement: time, cost, quality, and flexibility.											
... metrics that are chosen on a case-by-case basis according to specific IT projects and processes.				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... business driven metrics that are routinely used to measure performance and integrated into an assessment IS.		
The standardization of IT management activities ...											
... brings about bureaucracy and lacks of individuality.				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... increases the efficiency and quality of business processes.		

A. Questionnaire

Please specify your experiences with IT management.

7. Which, if any, of the following challenges have you experienced in the last 12 months, that might motivate your organization to implement IT management practices?	Yes	No
Pressure from business needs including time, frequency of changes, and cost	<input type="checkbox"/>	<input type="checkbox"/>
Business misunderstands the IT delivery	<input type="checkbox"/>	<input type="checkbox"/>
Poor business requirements (e.g., incomplete)	<input type="checkbox"/>	<input type="checkbox"/>
Organizational aspects, such as acceptance, responsibilities, and skills	<input type="checkbox"/>	<input type="checkbox"/>
Lack of stable processes (e.g., multiple ways to do similar things)	<input type="checkbox"/>	<input type="checkbox"/>
Speed of IT innovation	<input type="checkbox"/>	<input type="checkbox"/>
Operation/project costs	<input type="checkbox"/>	<input type="checkbox"/>
Lack of methods for IT management	<input type="checkbox"/>	<input type="checkbox"/>
Lack of system support	<input type="checkbox"/>	<input type="checkbox"/>
Disappointing results from automation	<input type="checkbox"/>	<input type="checkbox"/>
Other (<i>Please specify</i>)	<input type="checkbox"/>	<input type="checkbox"/>

A.2. Alignment of IT Management and Business Process Management

Please provide us with your overall evaluation of the alignment of IT management and business process management.

8. Please characterize the behavior of your IT department. We provide two contrasting situations from which you	2	1	0	1	2	can choose.
The underlying premise of our business process management is that ...						
... processes are documented if required by law.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... quality of products/services is determined by process quality.
... process models are useful.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... process modeling is a waste of time.
... there is little understanding of how processes are executed.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	... activities are executed w.r.t. process specifications; descriptions are update when needed.

A.2. Alignment of IT Management and Business Process Management

9. Which, if any, of the following benefits on business have you experienced after implementing IT management approaches?						Yes	No
Increase in transparency of processes						<input type="checkbox"/>	<input type="checkbox"/>
Clear ownership and responsibilities based on process orientation						<input type="checkbox"/>	<input type="checkbox"/>
Efficient process design						<input type="checkbox"/>	<input type="checkbox"/>
Higher degree of automation						<input type="checkbox"/>	<input type="checkbox"/>
Increase in efficiency of IT (e. g., reduce redundancy)						<input type="checkbox"/>	<input type="checkbox"/>
Better alignment between IT and business						<input type="checkbox"/>	<input type="checkbox"/>
More positive attitude toward process changes						<input type="checkbox"/>	<input type="checkbox"/>
Other (<i>Please specify</i>)						<input type="checkbox"/>	<input type="checkbox"/>

10. How would you rate the following to evaluate the performance of IT processes?	Not very important	Very im- portant	Don't know
Review of artifacts that are produced by performing a process.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Interviews with persons performing a process.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Interviews with persons who manage the process performance.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Quantitative data used to characterize the state of the department.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Quantitative data describing process performance.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Other (<i>Please specify</i>)	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>

11. Which, if any, of the following effects do IT improvements entail on business? (<i>Please mark all that apply</i>)	Stayed the same	Improved
Customer relationships (external/internal)	<input type="checkbox"/>	<input type="checkbox"/>
Product/service innovation/leadership	<input type="checkbox"/>	<input type="checkbox"/>
Cost of performing direct product/service tasks	<input type="checkbox"/>	<input type="checkbox"/>
Cost of help desks or other follow-on support functions	<input type="checkbox"/>	<input type="checkbox"/>
Time to get products/services into market/operations	<input type="checkbox"/>	<input type="checkbox"/>
Other (<i>Please specify</i>)	<input type="checkbox"/>	<input type="checkbox"/>

A. Questionnaire

A.3. Effectiveness of Reference Models

Reference models (frequently termed best practices) have become increasingly popular in process design and execution over the last 20 years; by reference models, we mean some sort of pattern, expressing the best way to treat a particular problem, which can be replicated in a similar situation or setting.

Please specify how aware you are with reference models.

12. Are you, yourself, or your organization aware with any kind of reference model? (Please mark all that apply)			
	Aware of existence	Aware of contents	No
You are personally	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Your company	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(If all No, ask Q. 17–19)			

13. Which of the following do you know, use, or are you considering to use within the next 12 months? (Please mark all that apply)				
	Aware	Consider to use	In use	Stopped
ITIL/ISO 20000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
COBIT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ISO 9000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Val IT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CMM/CMMI	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ISO17799/ISO 27000	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
COSO ERM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Internally developed framework	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other (Please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please specify how your IT department uses reference models.

14. Does your IT department use or consider to use reference models for any of the following activities? (Please mark all that apply)	
Do you focus on efficient means to perform and organize certain how-to aspects of management, such as service delivery?	<input type="checkbox"/>
Do you organize the processes in terms of responsibilities, controls ⁸⁵ , and organization?	<input type="checkbox"/>
Do you emphasize quality standards applied to specific domains (e.g., security)?	<input type="checkbox"/>
Do you focus on improving processes, performance or other, not focusing on how-to aspects?	<input type="checkbox"/>
Other (Please specify)	<input type="checkbox"/>

⁸⁵ Key performance indicators (KPIs) and key goal indicators (KGIs)

A.3. Effectiveness of Reference Models

15. To what extent would you disagree or agree that ...	Strongly disagree	Strongly agree	Don't know
...the compliance of IT processes with reference models needs to be measured and controlled?	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
...the use of reference models is integrated with your IT management?	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Other (<i>Please specify</i>)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>

16. Which of the following does your IT department use for maintaining reference models? (<i>Please mark all that apply</i>)
Audit <input type="checkbox"/>
Certification <input type="checkbox"/>
IT controlling <input type="checkbox"/>
Quality control <input type="checkbox"/>
Other (<i>Please specify</i>)

17. How does your IT department analyze or are you considering to analyze your as-is processes?	Consider to use	In use	Not at all
Workshop	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Interview	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External consultant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other (<i>Please specify</i>)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

18. How frequently does your IT department analyze ... in the past 12 months?	0	1-3	4-6	7-12	>12	Don't know
...if the processes are implemented as described	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
...if the responsibilities and controls are met	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
...if the quality standards are met	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
...if the processes are continuously improved	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>
Other (<i>Please specify</i>)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	<input type="checkbox"/>

A. Questionnaire

Please provide us with your overall evaluation of the effectiveness of reference models.

19. To what extent would you disagree or agree that ...	Strongly disagree	Strongly agree	Don't know
...you are satisfied with the parts of the reference model, which your organization uses?	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
...the content or structure of the reference model allows you to help implementing effective IT management?	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>

20. Which of the following do you hope to address using your selected reference model?	Not very important	Very important	Don't know
Unknown	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Business and IT alignment	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Business process improvement	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
IT process improvement	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Time/cost reduction for process modeling	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Organizational aspects	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Risk mitigation	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
IT compliance	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Standardization of processes	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>
Other (<i>Please specify</i>)	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>		<input type="checkbox"/>

A.4. Effectiveness of Process Mining

As you may know, some process discovery techniques are being developed using new scientific techniques. The general idea is called “process mining” and includes tools to automatically infer process models from the process knowledge that implicitly resides in some ISs, such as WFMSs, ERP systems, or CRM systems.

21. Have you heard or read about process mining? (<i>Please check one</i>)
Yes <input type="checkbox"/>
No <input type="checkbox"/> (Skip Q. 22)

A.4. Effectiveness of Process Mining

22. Which of the following do you know, use, or are you considering to use within the next 12 months? <i>(Please mark all that apply)</i>					
	Aware	Consider to use	In use	Stopped	
ProM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Futura Process Intelligence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Fluxicon	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Pallas Athena	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
OpenConnect	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ARIS Process Performance Manager	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Fujitsu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Iontas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Other <i>(Please specify)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
23. How helpful would IT be if process models can be automatically derived?					
Would not be very helpful to your company.				<input type="checkbox"/> (Skip Q. 24)	
Would be very helpful to your company.				<input type="checkbox"/>	
24. Are there any situations you can imagine process mining to be helpful? <i>(Please mark all that apply)</i>					
	IT/ITSM processes		Business processes		
Process analysis	<input type="checkbox"/>		<input type="checkbox"/>		
Process improvement	<input type="checkbox"/>		<input type="checkbox"/>		
Process compliance with reference models	<input type="checkbox"/>		<input type="checkbox"/>		
Other <i>(Please specify)</i>	<input type="checkbox"/>		<input type="checkbox"/>		
25. To what extent would you disagree or agree that the following barriers to an industrial application of process mining exist?					
	Not very important		Very important		Don't know
Unknown	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of skills and expertise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of sound commercial applications	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Organizational aspects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other <i>(Please specify)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

A. Questionnaire

Profile

Please provide us with details of your person and your occupation. The questions marked with a * are optional.

26. Please provide details of your person.	
What is your name?
What is your e-mail address? *
What is the highest number of years of school you completed?* Years
Graduated as:*
27. Please provide details of your occupation.	
For whom do you work?
What is your current role?
How many employees does your IT department employ in total? People
28. Are you interested in knowing more in detail about ...	
	<input type="checkbox"/> Yes <input type="checkbox"/> No
... reference models?	<input type="checkbox"/> <input type="checkbox"/>
... process mining?	<input type="checkbox"/> <input type="checkbox"/>
29. If you have any comments related to our questionnaire, please share them with us.	
.....	
.....	
.....	

Thank you for taking a moment to answer the questions.

B. Analysis

A discussion of the results of the questionnaire's analysis follows. Note that the responses to questions two, four, seven, nine, twelve, thirteen, seventeen, eighteen, twenty to twenty-two, and twenty-four have been already discussed in Section 5.3.1.

B.1. Importance and Performance of IT and IT Management

The survey evinced that the IT manager was predominantly responsible for IT management (see Figure B.1).

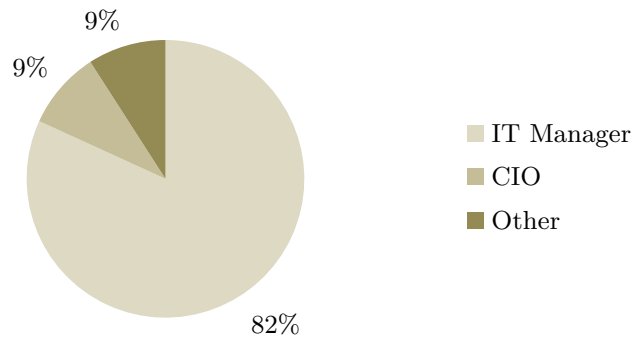


Figure B.1.: Responsibility for IT management

The responses indicated that less than a third committed their resources to process improvement and process communication with business, meaning that a clear majority (70.8 %) committed their resources to customary IT tasks (see Figure B.2).

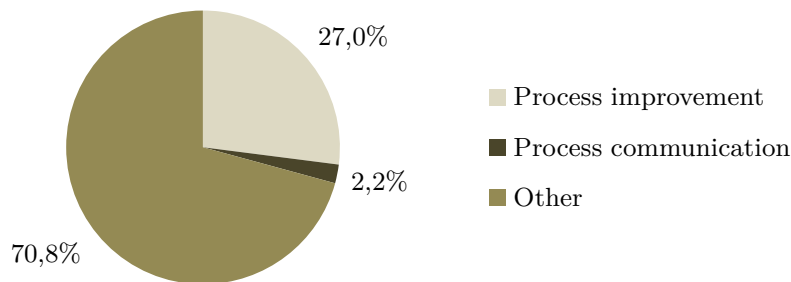


Figure B.2.: Resources committed to process improvement and process communication

B. Analysis

Invariably, respondents indicated that IT performance measurement is indispensable⁸⁶. All respondents deemed it important to measure operational performance, 73 % of them to measure the appropriate use of IT resources, 73 % to measure the degree of adoption of best practices, 82 % to measure either internal or external customer satisfaction (or both of them), 73 % to measure the IT process maturity, 82 % to measure the quality of IT products or services (or of both of them), and 64 % to measure the IT-related risk analysis (see Figure B.3).

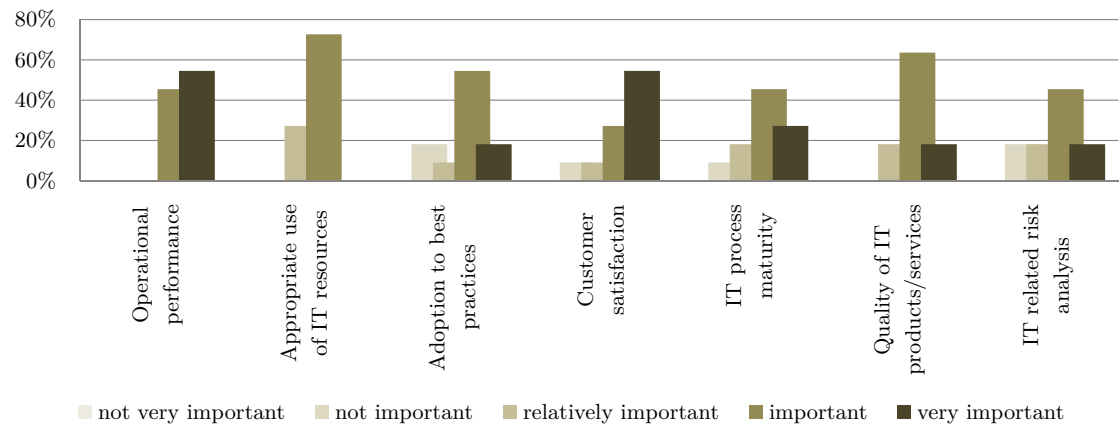


Figure B.3.: Subjects of measurement

Figure B.4 reflects the behavior of the IT department. Vis-a-vis the focus on IT, all respondents considered it customer-oriented rather than purely technology-oriented,

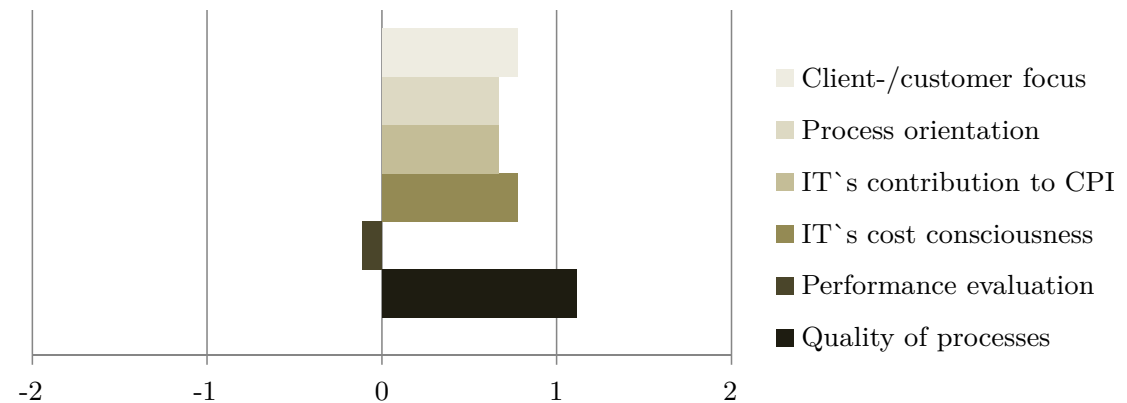


Figure B.4.: Behavior of the IT department vis-a-vis IT

⁸⁶ Note that only those who evaluated the measurement as important or very important are considered.

their organizational design process-oriented rather than function-oriented, and it as a means of achieving cost-efficient production of business products rather than achieving competitive advantage through its use. The respondents expected that IT was subject to a continuous improvement rather than meeting requirements according to specific IT projects. The respondents felt that their IT performance evaluation was based on ad hoc metrics rather than on business driven metrics that are routinely used to measure performance and integrated into an assessment IS. The respondents ascertained that the standardization of IT management activities increased the efficiency and quality of business processes rather than bureaucracy. Figure B.4 unfortunately reveals that IT management practices, albeit moving in the right direction, are still relatively low-to-moderate. This is woefully inadequate. Especially, there is a blatant inconsistency between the behavior toward process performance evaluation – a number of respondents ought to have been concerned with this evaluation vis-a-vis its attached importance (see Figure B.3), but clearly were not.

B.2. Alignment of IT Management and Business Process Management

Now the focus is on the overall evaluation of the alignment of IT management and BPM.

When asked about the underlying premise of business process management (see Figure B.5), the respondents indicated that the quality of their products or services, or of both of them, depended on process quality. They considered process models useful and process execution transparent vis-a-vis process specification. Although progress in process orientation is discernible, there is still substantial room for improvement.

In each of the following dimensions of Figure B.6, more than 30 % of those surveyed rated the importance as high or very high. Specifically, 36 % of respondents indicated that the review of artifacts produced by performing a process was an important means of measuring the performance of IT – 64 % thought so regarding interviews with per-

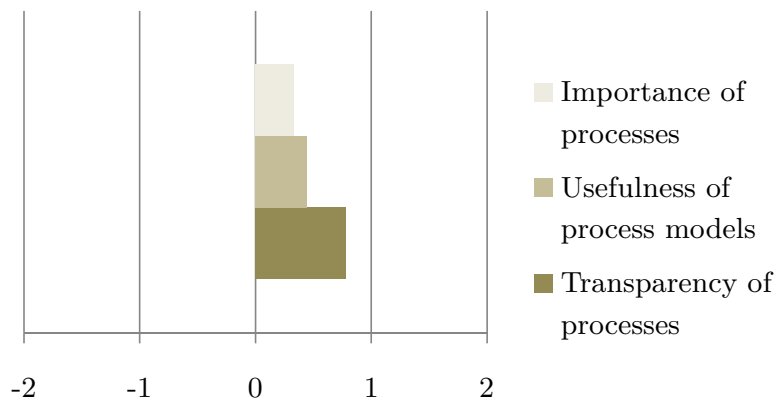


Figure B.5.: Behavior of the IT department vis-a-vis BPM

B. Analysis

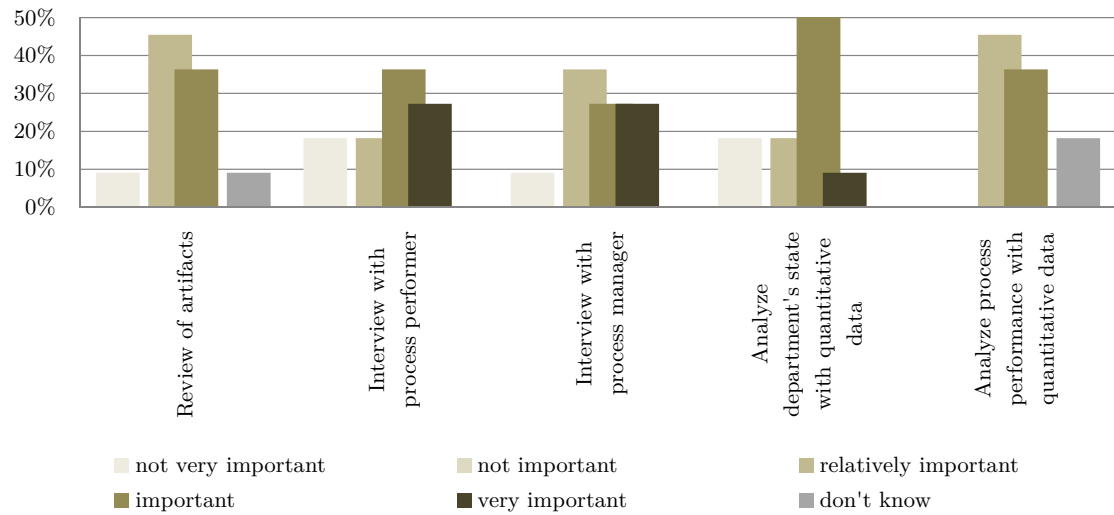


Figure B.6.: Means of performance evaluation

sons performing a process – 55 % thought so regarding interviews with persons who manage the process performance – 64 % thought so regarding quantitative data used to characterize the department's state – and 36 % thought so regarding quantitative data describing the process performance. Figure B.7 (multiple responses allowed) shows that nearly three-quarters experienced a positive business effect from IT improvements, that is, 72 % of the respondents said that both their customer relationships (external/internal)

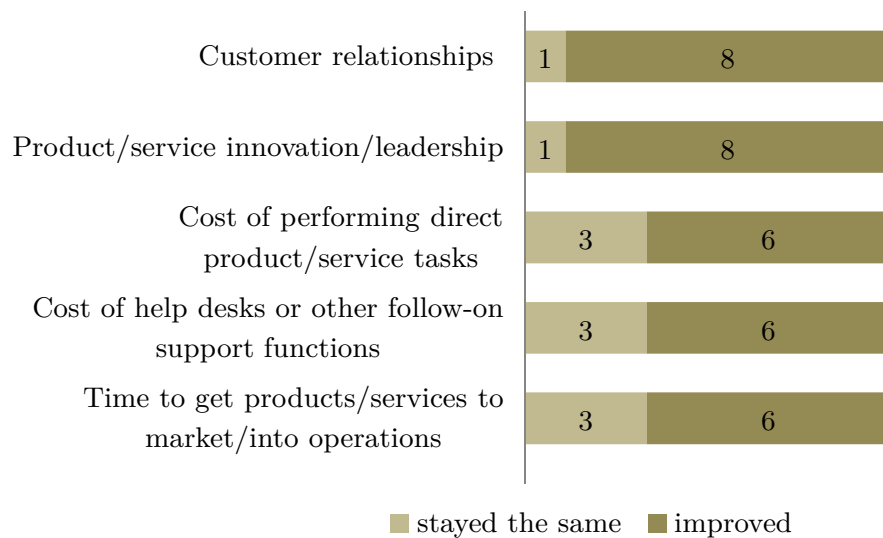


Figure B.7.: Effects of IT improvements on business

and product/service innovation/leadership had improved and 55 % noted improvements regarding cost of performing direct product/service tasks, cost of help desks or other follow-on support functions and time to get products/services into market/operations.

B.3. Effectiveness of Reference Models

As noted, a large increase in the adoption of reference models is evident, shown by the fact that 82 % of those surveyed said that they were aware of reference models and 55 % even knew their content (see Figure B.8). As to the organizations themselves, the numbers slightly decreased: 73 % were aware and 27 % knew the contents.

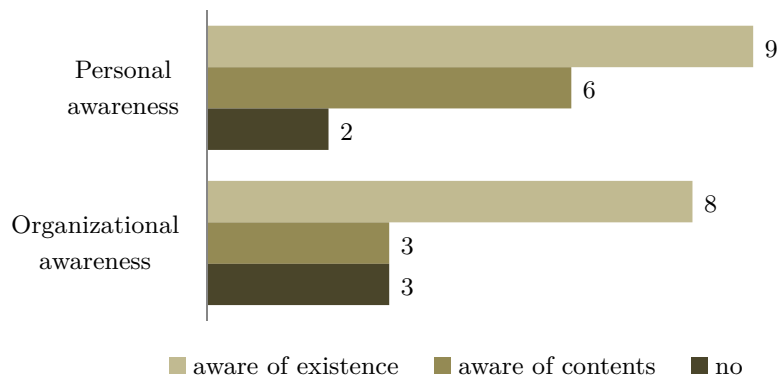


Figure B.8.: Awareness of reference models

Figure B.9 depicts that among the respondents ITIL and ISO 2000 are used most frequently (over 40 %) virtually the rest considered their use. Well-known reference

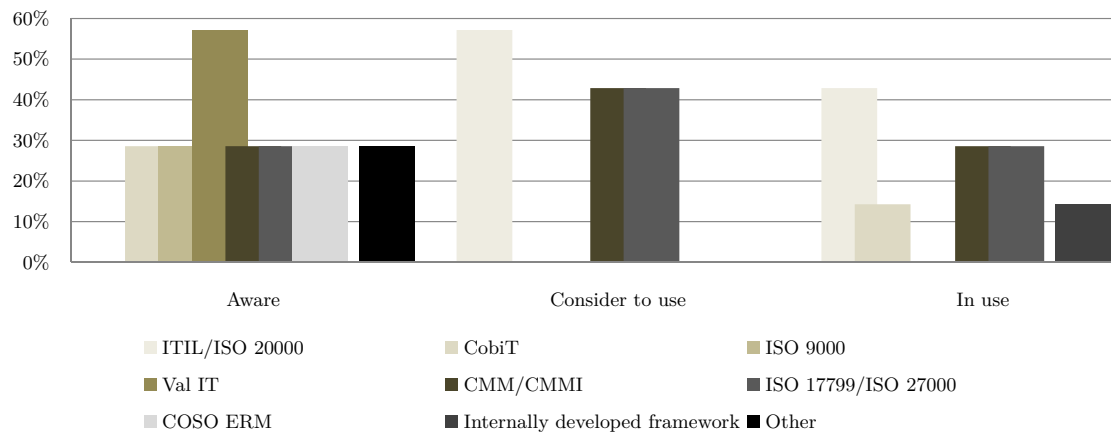


Figure B.9.: Use of reference models

B. Analysis

models include the ISO models, ITIL, CMMI, COBIT, Val IT, and The Committee of Sponsoring Organizations of the Treadway Commission Internal Control (COSO).

Almost 75 % of the respondents said that they organized the processes in terms of responsibilities and controls; 64 % emphasized quality standards (e. g., security); 45 % focused on efficient means to perform certain how-to aspects of management (e. g., service delivery); 45 % focused on improving processes, performance or otherwise, without focusing on how-to aspects (see Figure B.10, multiple responses allowed).

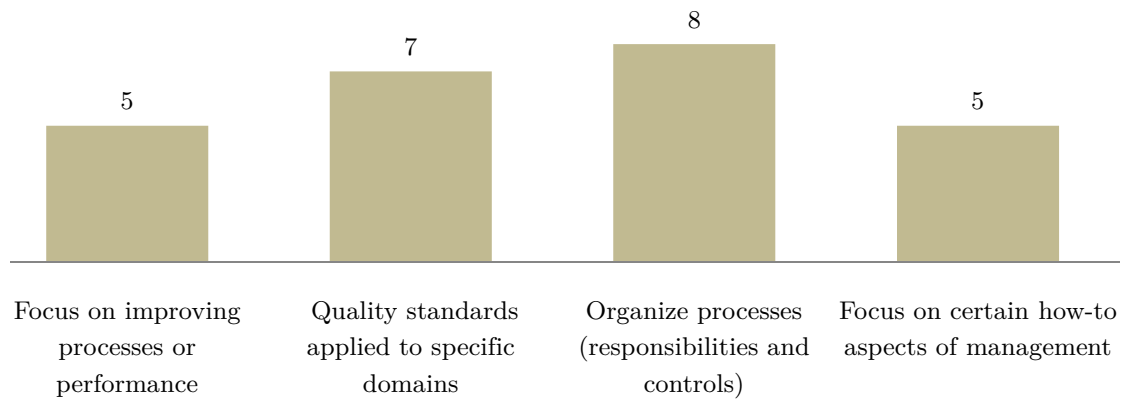


Figure B.10.: Purpose of reference models' use

Figure B.11 shows that more than 80 % of those surveyed asserted that the compliance of IT processes with reference models needed to be measured and controlled and 64 % asserted that reference models were intertwined with IT management.

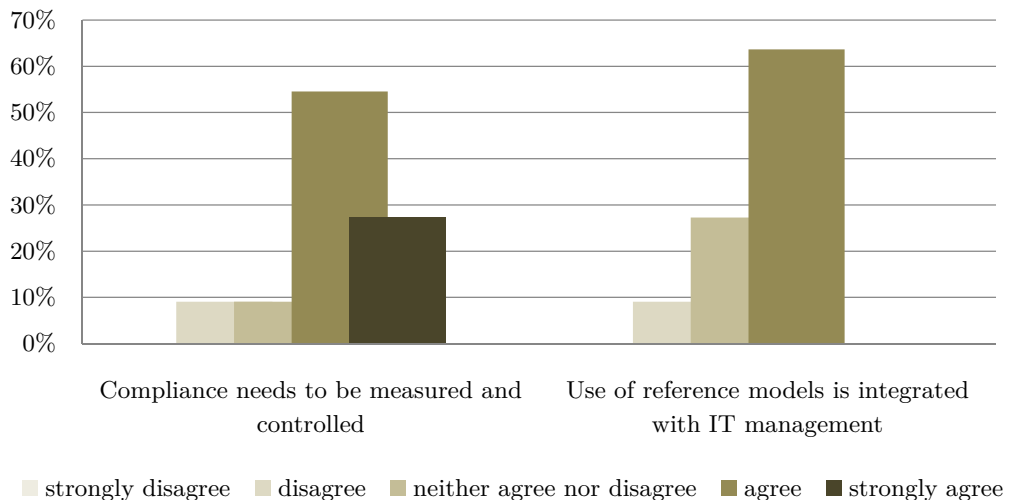


Figure B.11.: Characteristics of reference models

In those cases in which the respondents established the means to maintain the reference models, virtually all of these means were done manually, 55 % using audit, 36 % certification, 45 % IT controlling, and 55 % quality control (see Figure B.12, multiple responses allowed).

Despite the fact that only 45 % of respondents indicated that they were satisfied with the reference models, 73 % indicated that reference models contributed to effective IT management (see Figure B.13). This may be explained by the fact that the maintenance required very considerable effort – an effort the magnitude of which can be readily divined from Figure B.12 – as a result of which very considerable benefits were sought, with a degree of disappointment inevitable.

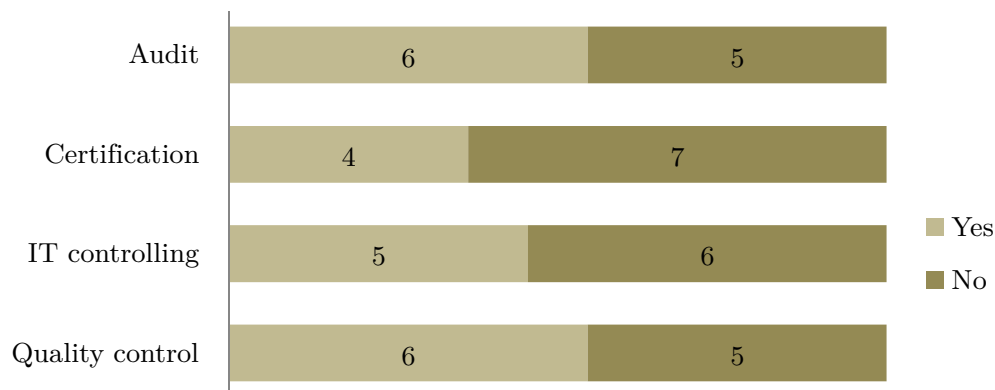


Figure B.12.: Means for maintaining reference models

B.4. Effectiveness of Process Mining

Surprisingly, a full 36 % of those surveyed knew process mining (see Figure B.14); 57 % of them were aware of ARIS PPM⁸⁷, 43 % of Pallas Athena, 29 % of Fujitsu, 14 % of ProM, 14 % of Futura Process Intelligence, and 14 % of other solutions (see Figure B.16).

None of the respondents used process mining. Those who were aware of ProM indicated that they would consider using it.

More than half of those surveyed believed that the automatic extraction of process models would aid in process modeling within their companies (see Figure B.15).

Invariably, respondents opined that the low diffusion in industry was due to obliviousness to process mining itself (100 %); 100 % cited lack of skills and expertise; 27 % cited lack of sound commercial applications; 73 % cited organizational aspects (see Figure B.17).

⁸⁷ Note that ARIS, Pallas Athena, and Fujitsu are well-known IT brands and therefore it may well be that the respondents ticked them without knowing that they were concerned with process mining.

B. Analysis

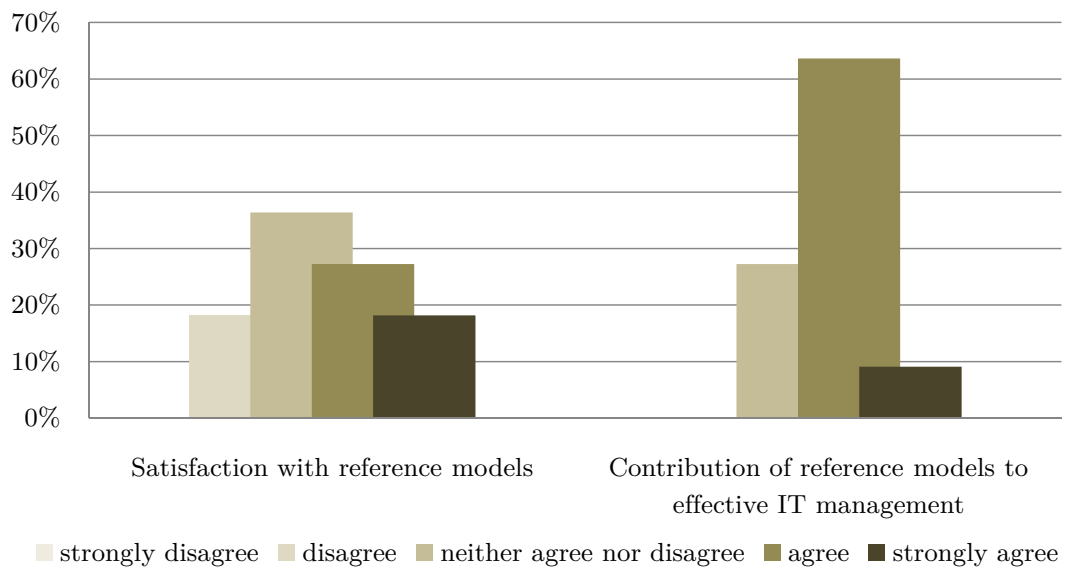


Figure B.13.: Experiences with reference models' use

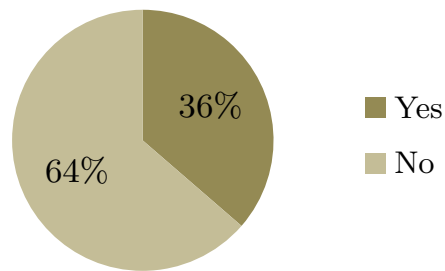


Figure B.14.: Awareness of process mining

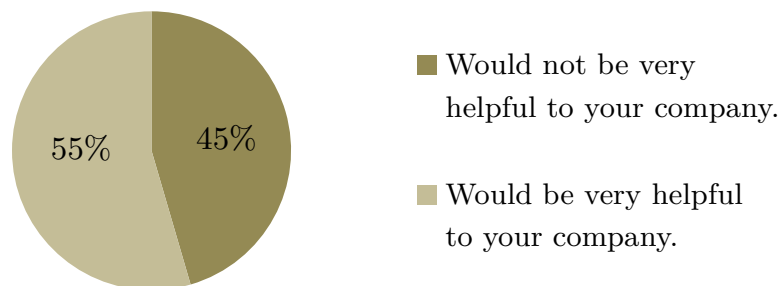


Figure B.15.: Helpfulness of automatic construction of process models

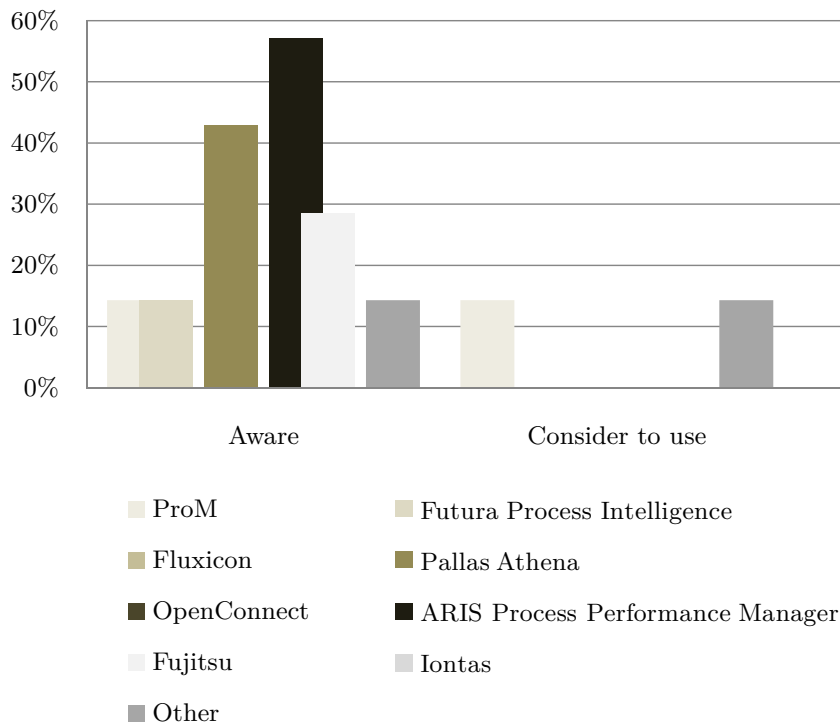


Figure B.16.: Awareness of process mining's products

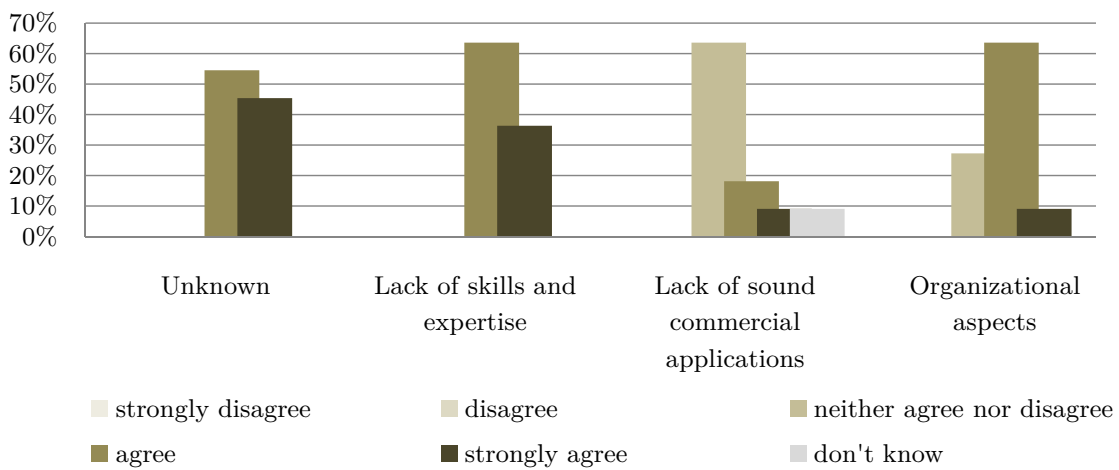


Figure B.17.: Barriers to industrial application of process mining

C. Process Models

This appendix embraces the process models of use cases Tel* that are not or partially shown in the preceding chapters.

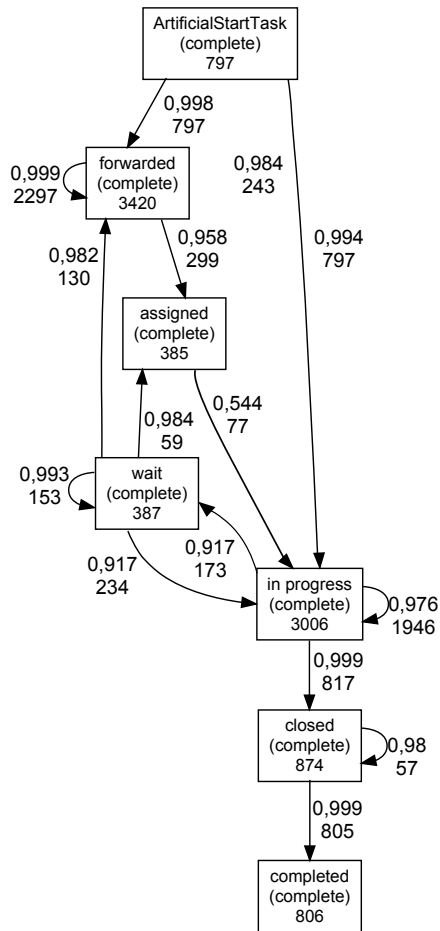


Figure C.1.: Process model Tel-1 of the transaction scheme

C. Process Models

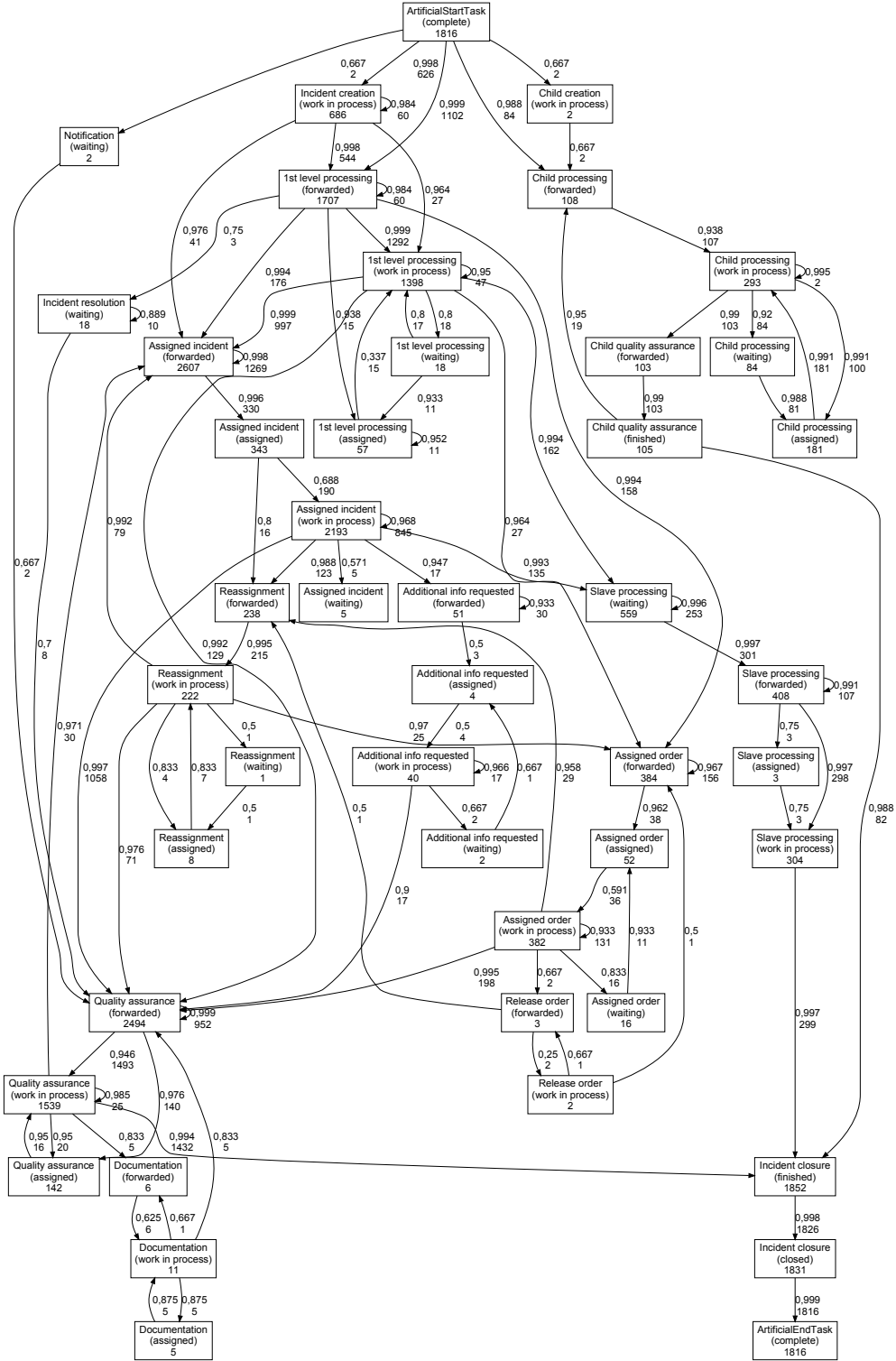


Figure C.2.: Process model Tel-2 of service S_1 on action level

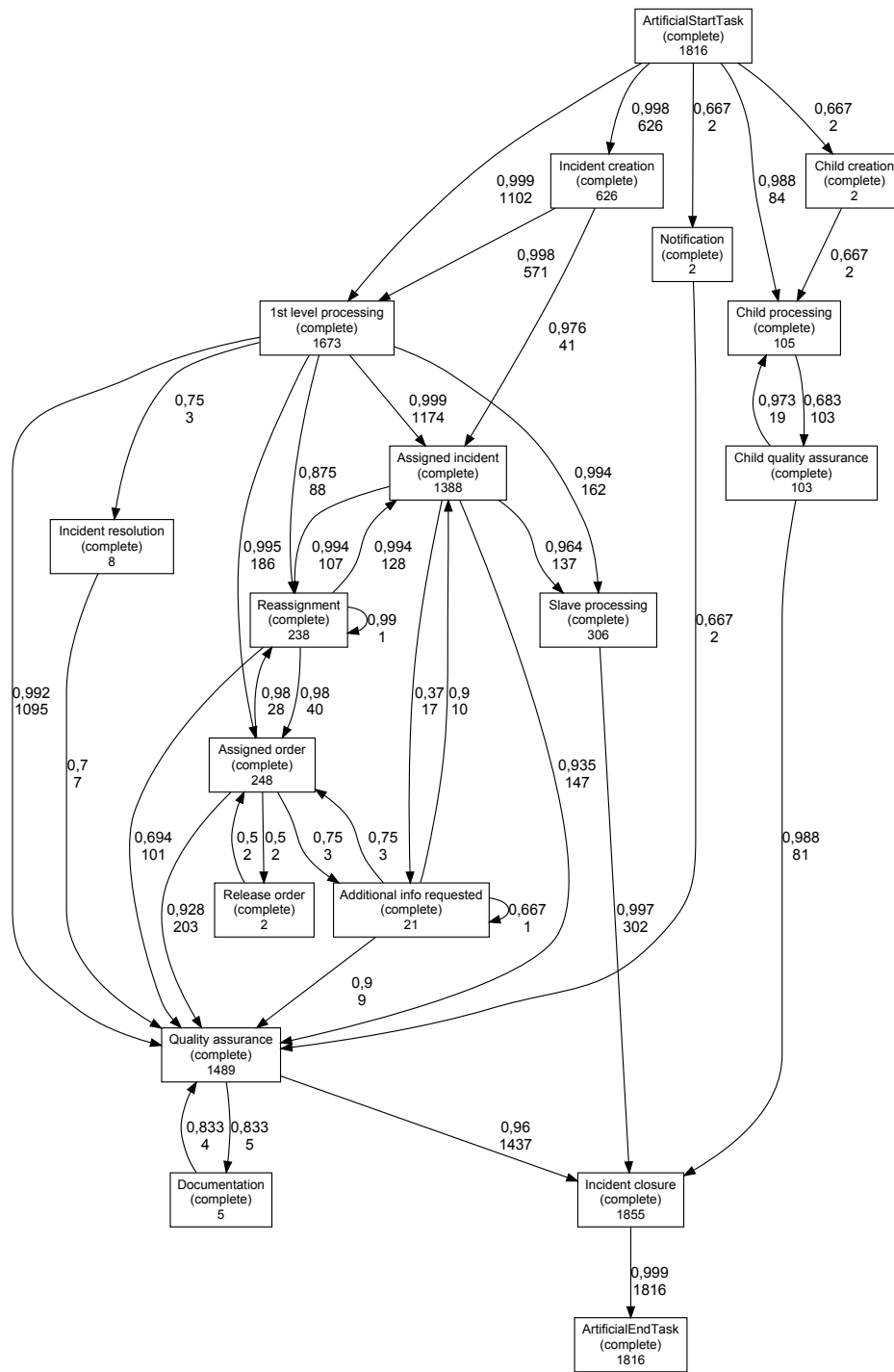


Figure C.3.: Process model Tel-2 of service S_1 on activity level

C. Process Models

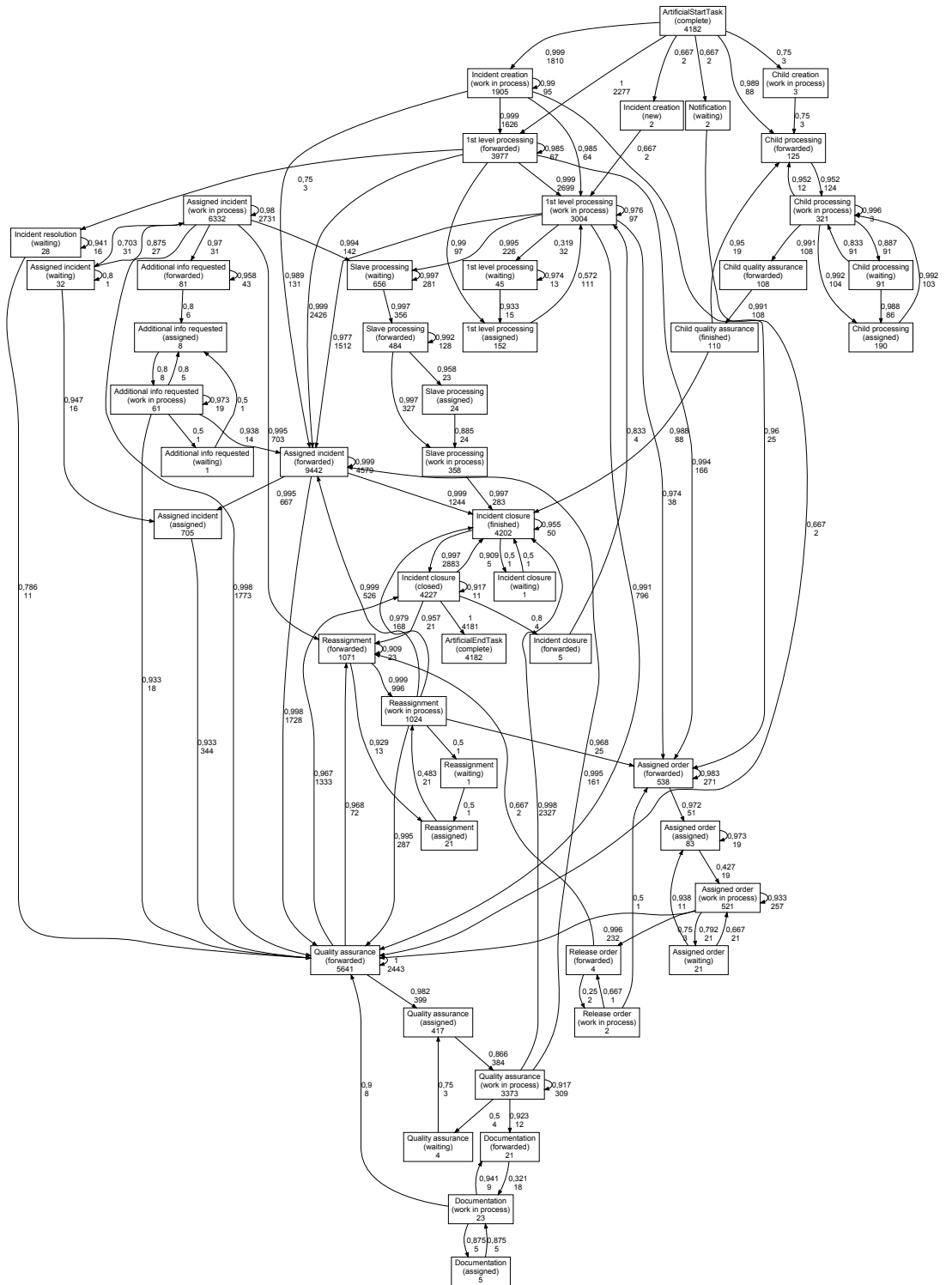


Figure C.4.: Process model Tel-3 of service S_2 on action level

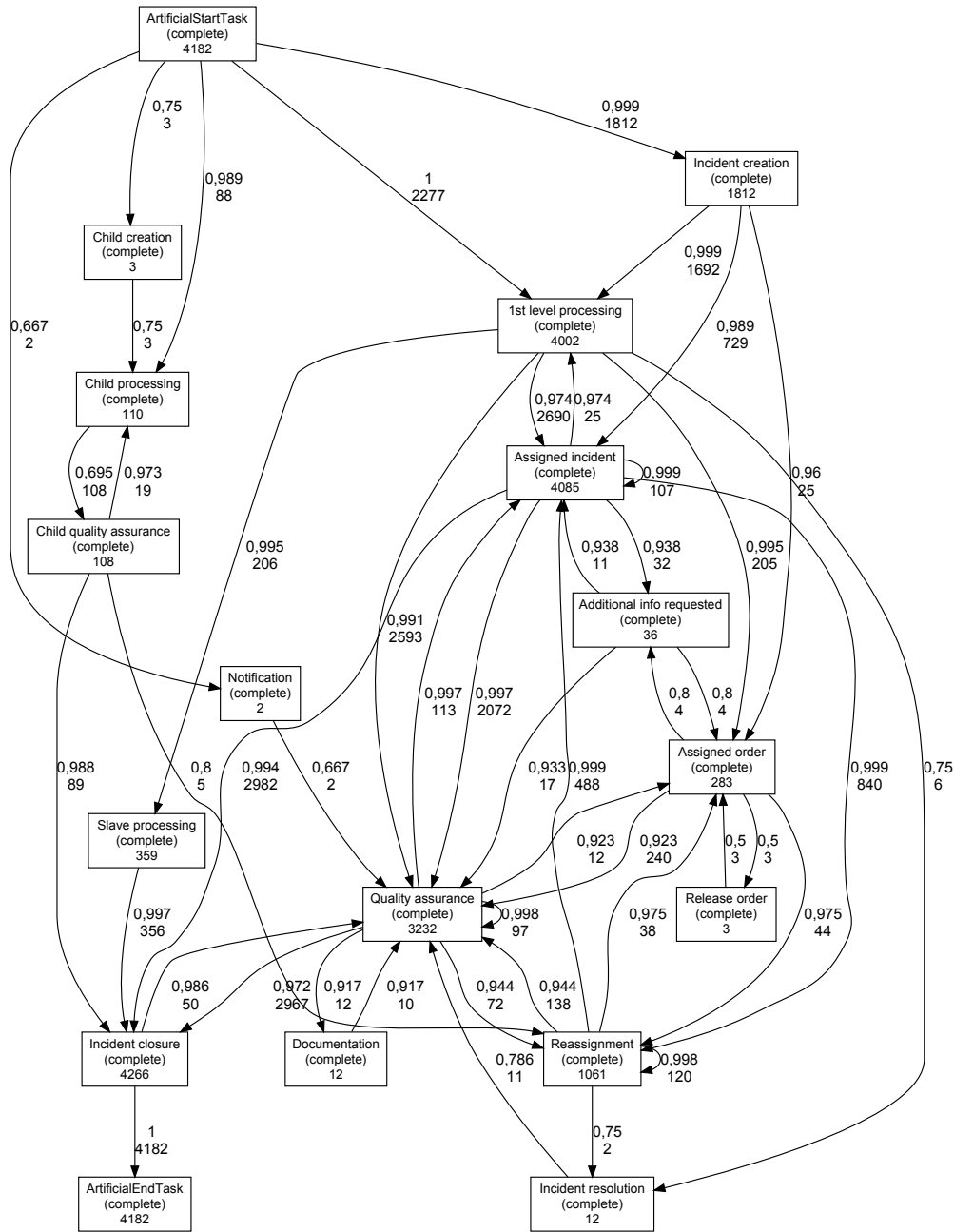


Figure C.5.: Process model Tel-3 of service S_2 on activity level

C. Process Models

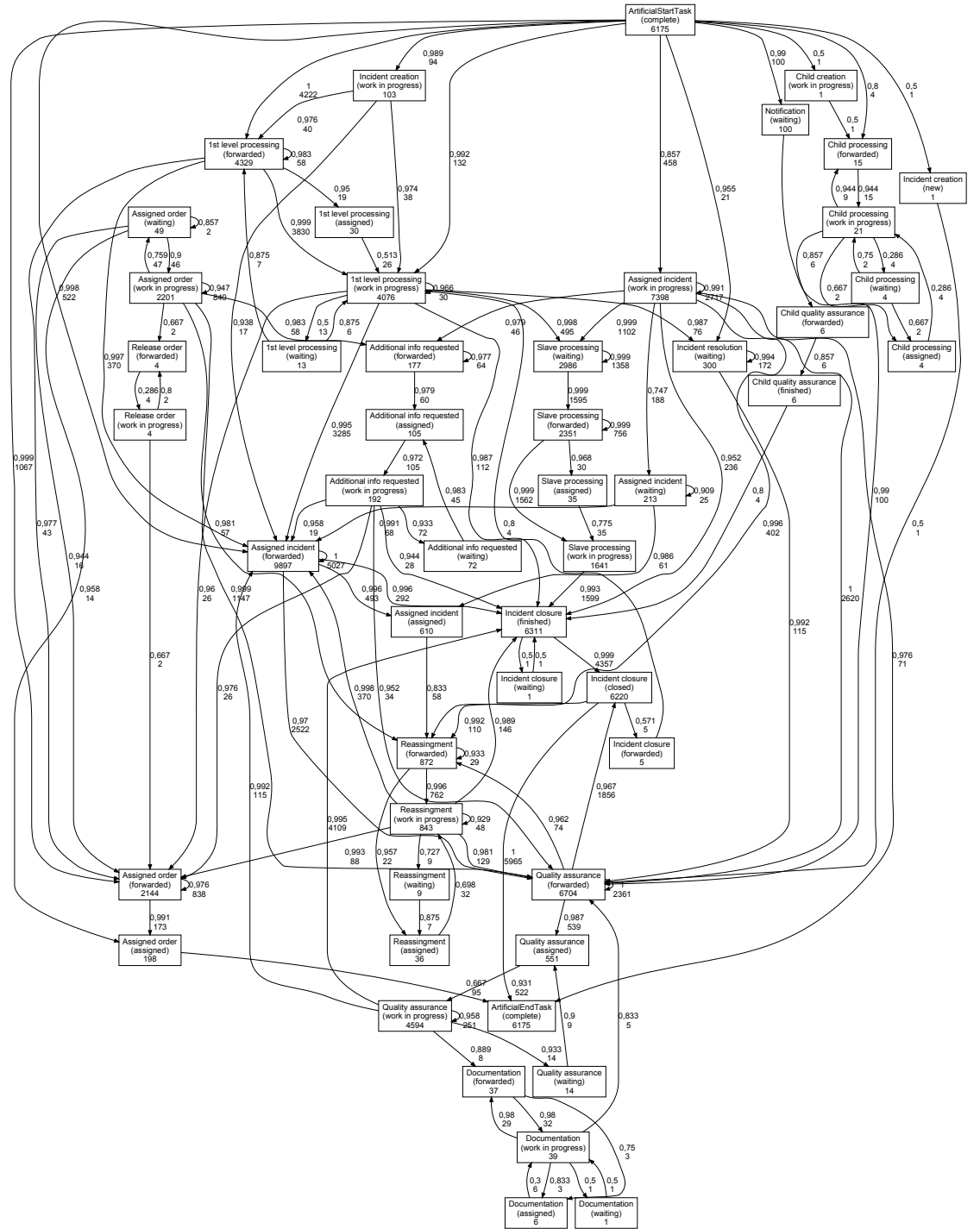


Figure C.6.: Process model Tel-4 of service S_3 on action level

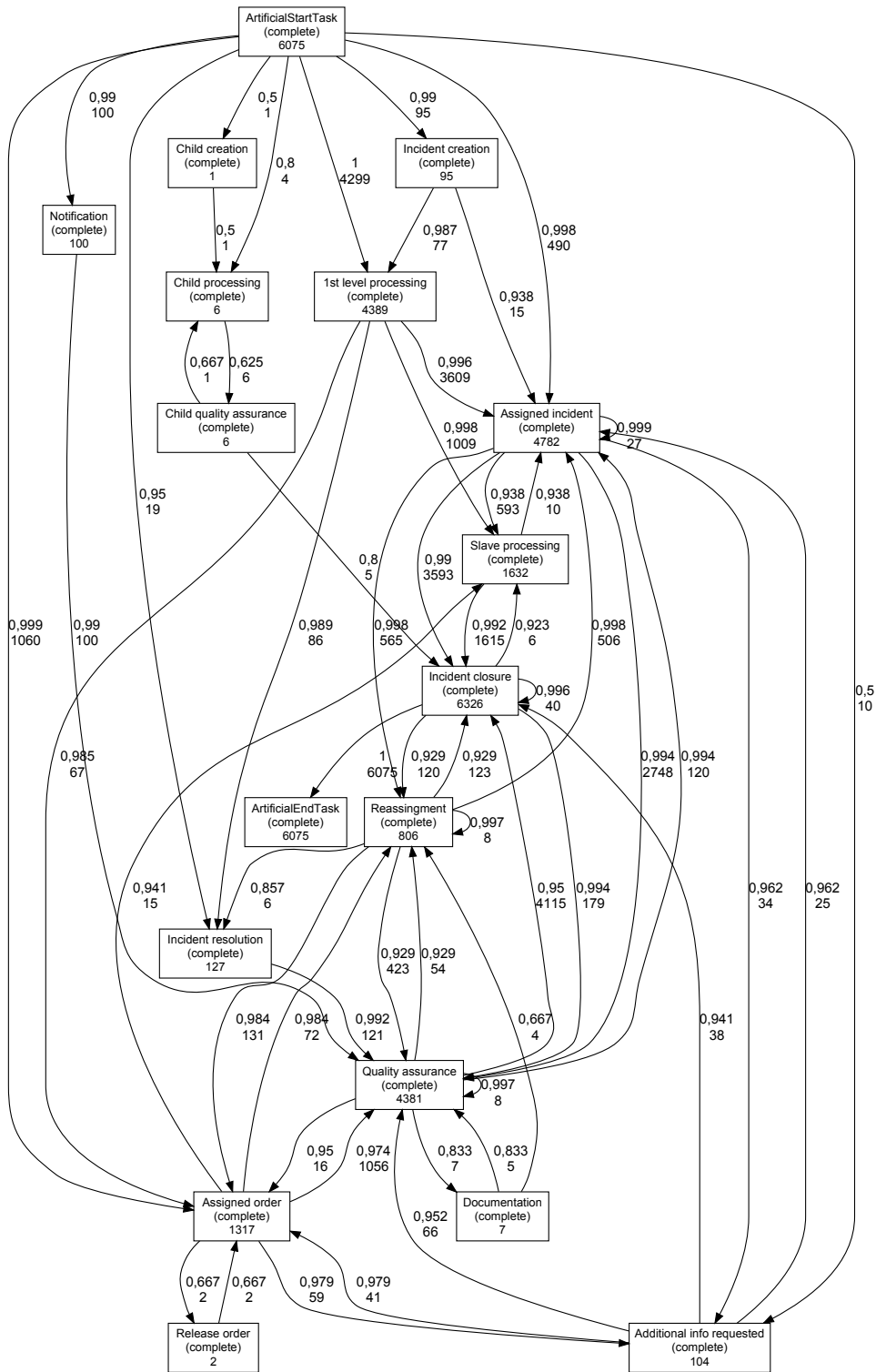


Figure C.7.: Process model Tel-4 of service S_3 on activity level

D. Process Mining Algorithms

This section addresses the process mining algorithms α -algorithm (van der Aalst et al., 2004) and *Heuristics miner* (Weijters et al., 2006).

The former is pertinent because it provides an introduction into process mining and some of its concepts are also used in the *Heuristics miner* – the algorithm that is applied to all the case studies herein. The starting point of the α -algorithm is an event log containing event traces. Each trace corresponds to an execution of a process. The ordering of events within a case is relevant, while the ordering of events among cases is of no importance (Dumas et al., 2005). Thus, an event log is defined as follows.

Definition 17 (Event Trace, Event Log)

Let T be a set of activities and T^* be a set of all traces of any length over T . $\sigma \in T^*$ is a sequence and $L : T^* \rightarrow \mathbb{N}$ is an event log. *

Any two activities in the log are classified according to one of the following ordering relations: $>_L$ (follows), \rightarrow_L (causal), \parallel_L (parallel), and $\#_L$ (unrelated). For example, if an activity always follows another activity, it is extremely likely that there is a causal dependency between both activities. These relations are extracted based on the MXML elements *WorkflowModelElement* and *ProcessInstance* and are defined as follows.

Definition 18 (Log-Based Ordering Relations)

Let L be an event log over T , that is, $L : T^* \rightarrow \mathbb{N}$. Let $a, b \in T$:

- (follows) $a >_L b$ iff there is a sequence $\sigma = t_1, t_2, t_3, \dots, t_n$ and $i \in \{1, \dots, n-1\}$ such that $\sigma \in L$ and $t_i = a$ and $t_{i+1} = b$,
- (causal) $a \rightarrow_L b$ iff $a >_L b$ and $b \not>_L a$,
- (parallel) $a \parallel_L b$ iff $a >_L b$ and $b >_L a$, and
- (unrelated) $a \#_L b$ iff $a \not>_L b$ and $b \not>_L a$. *

The *follows relation* describes the sequence in which activities are executed. The activity a is seen as a direct predecessor of activity b , if b is directly following a , namely $a >_L b$. The *causal relation* is derived from the *follows relation*. There is a causal dependency of a to b , if activity a is followed by activity b and the activity a never occurs after activity b , that is, $a \rightarrow_L b$. Two activities a and b can be executed in parallel, if they directly follow each other and can be executed in arbitrary order (i.e., $a \parallel_L b$ and $b \parallel_L a$). Two activities a and b are in no relation, if they never follow each other directly, that is, there is neither a direct causal relation nor a possible concurrency between the activities a and b .

Example 15 (Log-Based Ordering Relations)

Designating the activities *Receive order* as “A”, *Produce prod.* as “B”, *Reserve prod.* as “C”, *Ship order* as “D”, *Send invoice* as “E”, and *Close order* as “F”, for model (a) in Figure 2.2, a possible complete event log L is: $\{ ABEDF, ABDEF, ACEDF, ACDEF \}$.

The following ordering relations are inferred:

- (follows) $A >_L B, A >_L C, B >_L D, B >_L E, C >_L D, C >_L E, D >_L E, D >_L F, E >_L D$, and $E >_L F$.
- (causal) $A \rightarrow_L B, A \rightarrow_L C, B \rightarrow_L D, B \rightarrow_L E, C \rightarrow_L D, C \rightarrow_L E, D \rightarrow_L F$, and $E \rightarrow_L F$. Note that $D \nrightarrow_L E$ because $E >_L D$.
- (parallel) $D \parallel_L E$.
- (unrelated) $A \#_L F$. ◇

The α -algorithm uses these ordering relations to derive a WF-net. This special class of Petri nets is defined in detail in Sect. 6.4.1.

Definition 19 (α -Algorithm)

Let L be an event log over T . This is the algorithm’s definition.

1. $T_L = \{ t \in T \mid \exists \sigma \in L t \in \sigma \},$
2. $T_I = \{ t \in T \mid \exists \sigma \in L t = first(\sigma) \},$
3. $T_O = \{ t \in T \mid \exists \sigma \in L t = last(\sigma) \},$
4. $X_L = \{ (A, B) \mid A \subseteq T_L \wedge B \subseteq T_L \wedge \forall a \in A \forall b \in B a \rightarrow_L b \wedge \forall a_1, a_2 \in A a_1 \#_L a_2 \wedge \forall b_1, b_2 \in B b_1 \#_L b_2 \},$
5. $Y_L = \{ (A, B) \in X_L \mid \forall (A', B') \in X_L A \subseteq A' \wedge B \subseteq B' \implies (A, B) = (A', B') \},$
6. $P_L = \{ p_{(A, B)} \mid (A, B) \in Y_L \} \cup \{ i_L, o_L \},$
7. $F_L = \{ (a, p_{(A, B)}) \mid (A, B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A, B)}), b \mid (A, B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \},$ and
8. $\alpha(L) = (P_L, T_L, F_L).$ *

The algorithm works as follows. Step one determines the complete set of transitions T_L appearing in the event log L. Step two analyzes the set of initial transitions. Step three is the set of final transitions. Steps four and five define the places of the Petri net. Step four derives the causal relations, which allow for the correct mining of AND-splits and AND-joins as well as XOR-splits and XOR-joins constructs. Step five refines the set X_L by taking only the largest elements with respect to set inclusion, meaning that only

the minimal direct causal relations are considered. The places between the elements of Y_L , the source place i_L , and the sink place o_L are established in step six and connected to their respective input/output transitions in step seven. Finally, the Petri net in terms of a set of places P_L , a set of transitions T_L , and a set of edges F_L appears in step eight.

Example 16 (α -Algorithm)

Consider event log $L = \{ ABEDF, ABDEF, ACEDF, ACDEF \}$. $\alpha(L)$ is defined as follows.

1. $T_L = \{ A, B, D, E, F, C \},$
2. $T_I = \{ A \},$
3. $T_O = \{ F \},$
4. $X = \{ (\{ A \}, \{ B \}), (\{ A \}, \{ C \}), (\{ B \}, \{ D \}), (\{ B \}, \{ E \}),$
 $(\{ C \}, \{ D \}), (\{ C \}, \{ E \}), (\{ D \}, \{ F \}), (\{ E \}, \{ F \}),$
 $(\{ B, C \}, \{ D \}), (\{ B, C \}, \{ E \}), (\{ A \}, \{ B, C \}) \},$
5. $Y = \{ (\{ D \}, \{ F \}), (\{ E \}, \{ F \}), (\{ B, C \}, \{ D \}), (\{ B, C \}, \{ E \}),$
 $(\{ A \}, \{ B, C \}) \},$
6. $P_L = \{ i_L, o_L, p(\{ A \}, \{ B, C \}), p(\{ B, C \}, \{ D \}), p(\{ B, C \}, \{ E \}), p(\{ D \}, \{ F \}),$
 $p(\{ E \}, \{ F \}) \},$
7. $F_L = \{ (i_L, A), (A, p(\{ A \}, \{ B, C \})), (p(\{ A \}, \{ B, C \}), B), (p(\{ A \}, \{ B, C \}), C),$
 $(B, p(\{ B, C \}, \{ D \})), (p(\{ B, C \}, \{ D \}), D), (C, p(\{ B, C \}, \{ E \})),$
 $(p(\{ B, C \}, \{ E \}), E), (D, p(\{ D \}, \{ F \})), (p(\{ D \}, \{ F \}), F), (E, p(\{ E \}, \{ F \})),$
 $(p(\{ E \}, \{ F \}), F), (F, o_L) \}.$
8. $\alpha(L) = (P_L, T_L, F_L)$ is described in terms of a Petri net in model (a) of

Figure 2.2.

◇

The α -algorithm presupposes that the event log is complete with regard to the $>_L$ relations – if one activity succeeds another, the event log should consist of at least one example of this behavior. The α -algorithm presupposes also that the event log is noise free, that is, everything registered in the log is correct (Alves de Medeiros, 2006, p. 37). Note that even if the event log is complete and does not contain noise,

there are a number of constructs the α -algorithm cannot correctly rediscover: invisible activities (i.e., activities only used for routing purposes), duplicate activities (i.e., two nodes describing the same activity), non-free-choice constructs (i.e., a choice between two activities controlled by an earlier choice between two activities), and short loops (i.e., an activity executed multiple times in sequence) (Dumas et al., 2005, p. 246).

Since event logs in actual business settings are rarely complete and noise free, the algorithm *Heuristics miner* has been developed – it is robust as to log omissions and noise. This Heuristics approach merits discussion.

The idea is to take the frequency of the sequences into account. Thus, the *Heuristics miner* uses a frequency-based metric to calculate the certainty of a causal relation between two events a and b (notations $a \Rightarrow_L b$). The calculated \Rightarrow_L values between the events of an event log are used in a Heuristics search for the right relations between events (i.e., $a >_L b$, $a \#_L b$, or $a \parallel_L b$). Below, the \Rightarrow_L metric is defined, and an example follows.

Definition 20 (Dependency Measure)

Let L be an event log over T and $a, b \in T$. Then $|a >_L b|$ is the number of times $a >_L b$ occurs in L . The *dependency measure* \Rightarrow_L is:

$$a \Rightarrow_L b = \left(\frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} \right) \text{ if } (a \neq b) . \quad *$$

The more frequently an activity a directly follows another activity b , and the less frequently the opposite occurs, the higher the probability that a causally follows b . The $a \rightarrow_L b$ measure always yields a value between -1 and 1.

Example 17 (Dependency Measure)

1. Assuming that in ten traces activity A is directly followed by activity B but reverse occurs only once, then

$$A \Rightarrow_L B = \left(\frac{10 - 1}{10 + 1 + 1} \right) .$$

It follows from the value of $A \rightarrow_L B = .75$ that the causal relation is not completely reliable.

2. Assuming an event log with one hundred traces in which A is directly followed by B but reverse occurs only once, then

$$A \Rightarrow_L B = \left(\frac{100 - 1}{100 + 1 + 1} \right) .$$

Given the high value of $A \rightarrow_L B = .97$, one can conclude that the causal relation is quite reliable. \diamond

Now, the dependency/frequency metric is calculated – it consists of all \Rightarrow_L values for all possible activity combinations.

Example 18 (Frequency-Based Metric)

Consider the event log $L = \{ABEDF^9, ABDEF^9, ACEDF^9, ACDEF^9, ABCEDF, ADEF\}^{88}$. The 38 sequences consist of four traces, which appear nine times, and the two incorrect traces ABCEDF and ADEF. The frequency-based metric can be calculated as follows.

\Rightarrow_L	A	B	C	D	E	F
A	0.0	.95	.95	.50	0.0	0.0
B	-.95	0.0	.50	.90	.90	0.0
C	-.95	-.50	0.0	.90	.91	0.0
D	-.50	-.90	-.90	0.0	0.0	.95
E	0.0	-.90	-.91	0.0	0.0	.95
F	0.0	0.0	0.0	-.95	-.95	0.0

Like the α -algorithm, the *Heuristics miner* groups two activities according to the four ordering relations $>_L$, \rightarrow_L , \parallel_L , and $\#_L$. It additionally takes the frequency of the follows relations into account. The frequency is considered and there are two assumptions: (1) each non-initial activity must have at least one other activity that is its cause and (2) each non-final activity must have at least one dependent activity. The algorithm, therefore, takes the best candidates (with the highest $a \Rightarrow_L b$ score) from the dependency/frequency metric as the \rightarrow_L . Suitable threshold values compensate for infrequent behavior. From the basic relations the derived relations are inferred as follows.

Definition 21 (Log-Based Numerical Ordering Relations)

Let L be an event log over T , that is, $L : T^* \rightarrow \mathbb{N}$. Let $a, b \in T$:

- The basic relation is: (*follows*) $a > b$ iff $|a > b| \geq 0$.
- The derived relations are:

- (*causal*) $a \rightarrow_L b$ iff $|a >_L b| > 0$ and $|b >_L a| = 0$,
- (*parallel*) $a \parallel_L b$ iff $|a >_L b| > 0$ and $|b >_L a| > 0$, and
- (*unrelated*) $a \#_L b$ iff $|a >_L b| = 0$ and $|b >_L a| = 0$. *

Example 19 (Log-Based Numerical Ordering Relations)

The best candidates in the dependency/frequency metric are highlighted. All having a dependency threshold of .90 or higher were chosen.

⁸⁸ The upper case x denotes that this sequence occurs x-times in the event log, for example, $\{ABEDF^9\}$ occurs nine times.

D. Process Mining Algorithms

- Basic relations: $A > B, A > C, B > D, B > E, C > D, C > E, D > E, D > F,$ and $E > F$.
- Derived relations:
 - (causal) $A \rightarrow_L B, A \rightarrow_L C, B \rightarrow_L D, B \rightarrow_L E, C \rightarrow_L D, C \rightarrow_L E, D \rightarrow_L F,$ and $E \rightarrow_L F$.
 - (parallel) $D \parallel_L E$.
 - (unrelated) $A \#_L F$. ◇

Based on the derived ordering relations a dependency/frequency graph is constructed. The next step is the calculation of the type of splits and joins. The underlying representation is not a Petri net but a so-called Heuristics net (see Figure D.1).

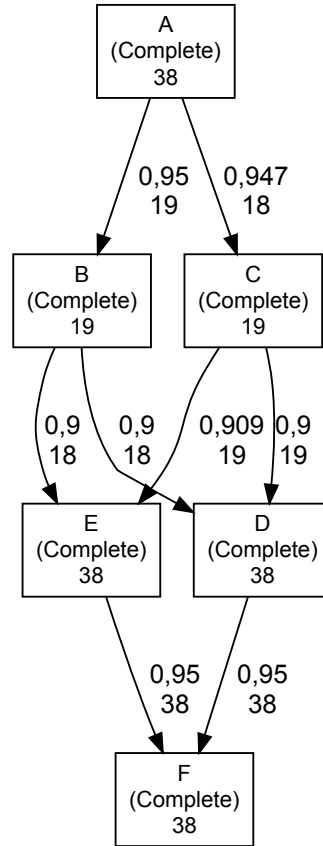


Figure D.1.: Example Heuristics net

E. SCOR D2 Process Strategy

The SCOR (D2) scenario is applied in one use case, illustrated herein. The (D2) scenario represents the delivery strategy in which products are delivered only in response to a customer order (see Figure E.1).

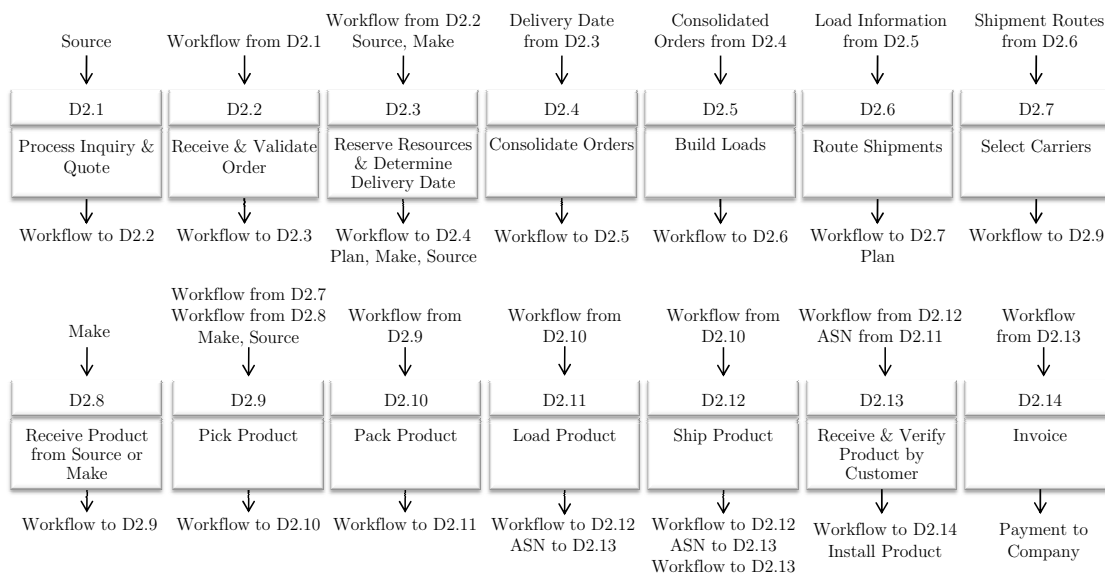


Figure E.1.: SCOR process category D2

Though, the process may begin with the receipt of a concrete customer order for one or more items, it also may begin with earlier inquiries or a requests for quotes (D2.1 Process Inquiry & Quote). After the receipt of a customer order it is entered into an order processing system for validation (D2.2 Receive & Validate Order), which triggers the reservation of required inventory and planned capacity for specific orders. Then a delivery date is scheduled (D2.3 Reserve Resources & Determine Delivery Date). If procurable, orders are grouped for least cost or best service fulfillment (D2.4 Consolidate Orders). Transportation modes are then selected and efficient loads are built (D2.5 Build Loads). In the next step (D2.6 Route Shipments), the loads are consolidated and routed by mode and location. Afterwards a carrier is selected (D2.7 Select Carriers).

After the products are received, say, by a manufacturer, that has received raw material, these products are verified and stored (D2.8 Receive Product from Source or Make). The pick process contains a series of activities including retrieving orders to pick, verifying

E. SCOR D2 Process Strategy

inventory availability, building the pick wave, picking the product, recording the pick, and delivering product to packing area in response to the order (D2.9 Pick Product). Then after things, such as sorting or packing, and pasting tags, the picked products are delivered to the shipping area for loading (D2.10 Pack Product), where they are placed onto modes of transportation (D2.11 Load Product). The loading process includes the generation of the documentation (e. g., advanced shipping notification (ASN)) necessary to meet internal, customer, carrier, and government requirements and triggers the process of shipping the product to the customer's site. The process of shipping (D2.12 Ship Product) comprises the carrier selection and a series of activities including loading the product onto modes of transportation and generating the documentation. Finally, the customer receives the shipment and verifies it for completeness and delivery terms (D2.13 Receive & Verify Product by Customer). The process ends with a signal, which is sent to the financial department that the order has been shipped and that the billing process should begin (D2.14 Invoice).

Bibliography

- Acrys Consulting. Ergebnisse der Studienumfrage Status quo Geschäftsprozessmanagement 2005, Retrieved 16 September 2010, from www.acrys.com/de/PDF/GPM_Umfrageergebnisse_2005.pdf, 2005.
- Rob Addy. *Effective IT service management - to ITIL and beyond*. Springer, Berlin, Heidelberg, Germany, 2007.
- Rakesh Agrawal, Dimitrios Gunopulos, and Frank Leymann. Mining process models from workflow logs. In Hans-Jürg Schenk, Felix Saltor, Isidro Ramos, and Gustavo Alonso, editors, *Proceedings of the 6th International Conference on Extending Database Technology*, pages 469–483, London, UK, 1998. Springer.
- Thomas Allweyer and August-Wilhelm Scheer. Modellierung und Gestaltung adaptiver Geschäftsprozesse. In August-Wilhelm Scheer, editor, *Publication of the University of the Saarland, Issue 115*, pages 1–26, Saarbrücken, Germany, 1995.
- Ana Karla Alves de Medeiros. *Genetic Process Mining*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2006.
- Ana Karla Alves de Medeiros, Boudewijn F. van Dongen, Wil M. P. van der Aalst, and A. J. M. M. Weijters. Process mining: Extending the α -algorithm to mine short loops. Technical report, Eindhoven University of Technology, Eindhoven, The Netherlands, 2004.
- Australian BPM Community of Practice. Proposed definition of business process management, retrieved 01 September 2009, from <http://www.bpm-roundtable.com>, 2009.
- D. Avison, J. Jones, P. Powell, and D. Wilson. Using and validating the strategic alignment model. *Journal of Strategic Information Systems*, 13(3):223–246, 2004.
- Tom Olavi Bangemann. *Shared services in finance and accounting*. Gower Publishing Limited, Hampshire, England, 2005.
- A.P. Barros, G. Decker, M. Dumas, and F. Weber. Correlation patterns in service-oriented architectures. In M.B. Dwyer and A. Lopes, editors, *Proceedings of the 10th International Conference on Theory and Practice of Software*, volume 4422 of *Lecture Notes in Computer Science*, pages 245–259, Berlin, Heidelberg, Germany, 2007. Springer.

Bibliography

- Anitesh Barua, Charles H. Kriebel, and Tridas Mukhopadhyay. Information technologies and business value: An analytic and empirical investigation. *Information Systems Research*, 6(1):3–23, 1995.
- Basel Committee on Banking Supervision (BCBS). Basel II: International convergence of capital measurement and capital standards. Retrieved 18 October 2010, from www.bis.org/publ/bcbs128.pdf, 2006.
- T. Basten and Wil M. P. van der Aalst. Inheritance of behavior. *Journal of Logic and Algebraic Programming*, 47(2):47–145, March/April 2001 2001.
- George W. Beeler, Jr. and Dana Gardner. A requirements primer. *Queue*, 4(7):22–26, 2006. ISSN 1542-7730.
- Izak Benbasat, David K. Goldstein, and Melissa Mead. The case research strategy in studies of information systems. *MIS Quarterly*, 11(3):369–386, 1987. ISSN 0276-7783.
- L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. In *Proceedings of the 7th IEEE International Symposium on String Processing Information Retrieval*, pages 39–48, Washington, DC, USA, 2000. IEEE Computer Society. ISBN 0-7695-0746-8.
- Peter Brooks. *Metrics for IT service management*. Van Haren Publishing, Zaltbommel, The Netherlands, 2006.
- E. Brynjolfsson and L. Hitt. Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science*, 42(4):541–558, 1996.
- Erik Brynjolfsson. The productivity paradox of information technology. *Communications of the ACM*, 36:66–77, 1993.
- Peter Buchholz and Peter Kemper. Hierarchical reachability graph generation for Petri nets. *IEEE Transactions on Software Engineering, IEEE Press*, 21(3):281–315, 2002. ISSN 0925-9856.
- Jacques Bughin, Michael Chui, and James Manyika. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *MIS Quarterly*, 3(1):1–14, 2010.
- Jorge Cardoso. Business process quality metrics: Log-based complexity of workflow patterns. In *Proceedings of the 15th International Conference on the Move to Meaningful Internet Systems*, volume 4803 of *Lecture Notes in Computer Science*, pages 427–434, Berlin, Heidelberg, Germany, 2007. Springer.
- N. G. Carr. *Does IT matter?* Harvard Business School Press, Boston, MA, USA, 2004.
- Fabio Casati, Malu Castellanos, Umeshwar Dayal, and Norman Salazar. A generic solution for warehousing business process data. In *Proceedings of the 33rd International Conference on Very Large Data Bases*, pages 1128–1137, Danvers, MA, USA, 2007. VLDB Endowment, Inc. ISBN 978-1-59593-649-3.

- Y. E. Chan. Why haven't we mastered alignment? The importance of the informal organizational structure. *MIS Quarterly Executive*, 1(2):97–112, 2002.
- Y. E. Chan, S. L. Huff, D. W. Barclay, and D. G. Copeland. Business strategy orientation, information systems orientation, and strategic alignment. *Information Systems Research*, 8(2):125–150, 1997.
- Eliot D. Chapple and Leonard R. Sayles. *The measure of management: Designing organizations for human effectiveness*. Macmillan, New York, NY, USA, 1961.
- Peter Pin-Shan Chen. The entity-relationship model - toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36, 1976. ISSN 0362-5915.
- R. Chen and August-Wilhelm Scheer. Modellierung von Prozessketten mittels Petri-Netz-Theorie. In August-Wilhelm Scheer, editor, *Publication of the University of the Saarland, Issue 107*, Saarbrücken, Germany, 1994.
- Committee of Sponsoring Organizations of the Treadway Commission (COSO). *Enterprise risk management - integrated framework*. American Institute of Certified Public Accountants, Jersey City, NJ, USA, 2004.
- Jonathan E. Cook and Alexander L. Wolf. Automating process discovery through event-data analysis. In *Proceedings of the 17th ACM International Conference on Software Engineering*, pages 73–82, New York, NY, USA, 1995. ACM.
- Jonathan E. Cook and Alexander L. Wolf. Event-based detection of concurrency. In *Proceedings of the 6th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 35–45, New York, NY, USA, 1998. ACM.
- Jonathan E. Cook and Alexander L. Wolf. Software process validation: Quantitatively measuring the correspondence of a process to a model. *ACM Transactions on Software Engineering and Methodology*, 8(2):147–176, 1999.
- Jonathan E. Cook, Cha He, and Changjun Ma. Measuring behavioral correspondence to a timed concurrent model. In *Proceedings of the 17th IEEE International Conference on Software Maintenance*, pages 332–350, Washington, DC, USA, 2001. IEEE Computer Society. ISBN 0-7695-1189-9.
- Cross Industry Standard Process for Data Mining. Retrieved 05 May 2008, from <http://www.crisp-dm.org>, 2008.
- T. Curran, G. Keller, and A. Ladd. *SAP R/3 business blueprint: Understanding the business process reference model*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1997.
- Thomas H. Davenport. *Process innovation. Reengineering work through information technology*. Harvard Business School Press, Boston, MA, USA, 1993.

Bibliography

- G.B. Davis and S. Hamilton. *Managing Information*. Business One Irwin, Homewood, IL, USA, 1993.
- U. Dayal, M. Hsu, and R. Ladin. Business process coordination: State of the art, trends, and open issues. In Peter M. G. Apers, Paolo Atzeni, et al., editors, *Proceedings of the 27th International Conference on Very Large Data Bases*, pages 3–13, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers, Inc.
- H. de Beer. The LTL checker plugins: A reference manual. *Eindhoven University of Technology, Eindhoven, The Netherlands*, 2004.
- G. Decker and J. Mendling. Instantiation semantics for process models. In M. Dumas, M. Reichert, and M.-C. Shan, editors, *Proceedings of the 6th International Conference on Business Process Management*, volume 5240 of *Lecture Notes in Computer Science*, pages 164–179, Berlin, Heidelberg, Germany, 2008. Springer.
- J. Dehnert and Wil M. P. van der Aalst. Bridging the gap between business models and workflow specifications. *International Journal of Cooperative Information Systems*, 13(3):289–332, 2004.
- W. Edwards Deming. *Out of the crisis*. MIT Press, Cambridge, UK, 1982.
- Remco Dijkman. A classification of differences between similar business processes. In *Proceedings of the 11th IEEE International Conference on Enterprise Distributed Object Computing*, pages 37–47, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 0-7695-2891-0.
- Remco Dijkman. Diagnosing differences between business process models. In *Proceedings of the 6th International Conference on Business Process Management*, volume 5240 of *Lecture Notes in Computer Science*, pages 261–277, Berlin, Heidelberg, Germany, 2008. Springer.
- D. Dörner. Ändert das KonTraG die Anforderungen an den Abschlußprüfer? *Der Betrieb*, 51(1):1–8, 1998.
- Marlon Dumas, Wil M. P. van der Aalst, and A. H. M. ter Hofstede. *Process-aware information systems: Bridging people and software through process technology*. John Wiley & Sons, Inc., New York, NY, USA, 2005.
- Schahram Dustdar and Robert Gombotz. Discovering web service workflows using web services interaction mining. *International Journal of Business Process Integration and Management*, 1(4):256–266, 2007.
- Johann Eder, Georg E. Olivotto, and Wolfgang Gruber. A data warehouse for workflow logs. In *Proceedings of the 1st International Conference on Engineering and Deployment of Cooperative Information Systems*, volume 2480 of *Lecture Notes in Computer Science*, pages 1–15, Berlin, Heidelberg, Germany, 2002. Springer.

- Marc Ehrig, Agnes Koschmider, and Andreas Oberweis. Measuring similarity between semantic business process models. In *Proceedings of the 4th International Conference on Conceptual Modeling*, pages 71–80, Darlinghurst, Australia, 2007. Australian Computer Society, Inc.
- EPCglobal, Inc. EPC Information Services (EPCIS), specification, version 1.0.1. Retrieved 03 November 2007, from <http://www.epcglobalinc.org/standards/epcis>, 2007.
- Henri Fayol. *Administration industrielle et générale. Prévoyance, organisation, commandement, coordination, control*. Dunod, Paris, France, 1966.
- Norman E. Fenton and Shari Lawrence Pfleeger. *Software metrics: A rigorous and practical approach*. PWS Publishing Co., Boston, MA, USA, 1996.
- fluxicon. Retrieved 01 April 2010, from <http://www.fluxicon.com>, 2010.
- B. Flyvbjerg. Five misunderstandings about case study research. *Qualitative Inquiry*, 12(2):219–245, 2006.
- E.N. Fong and D.K. Jefferson. Reference models for standardization. *IEEE Computer Standards Conference*, 5(2):93–98, 1986.
- Henry Ford. *Today and tomorrow*. Doubleday, Page and Company, Garden City, NY, USA, 1926.
- Heinz Frank and Johann Eder. Towards an automatic integration of statecharts. In *Proceedings of the 18th International Conference on Conceptual Modeling*, volume 1728 of *Lecture Notes in Computer Science*, pages 430–444, Berlin, Heidelberg, Germany, 1999. Springer.
- Michael Füchle and Matthias E. Zierke. *SAP CRM Web Client - customizing and development*. Galileo Press, Bonn, Germany, 2009.
- Futura Reflect. Retrieved 01 March 2009, from <http://www.futuratech.nl>, 2009.
- Stuart D. Galup, Ronald Dattero, Jim J. Quan, and Sue Conger. An overview of IT service management. *Communications of the ACM*, 52(5):124–127, 2009. ISSN 0001-0782.
- Inc. Gartner. Gartner it infrastructure, operations & management, summit 2009 – post event brief. Retrieved 16 November 2010, from http://www.gartner.com/it/content/750200/750233/iom_2009_trip_report.pdf, 2009.
- Gartner Research. Gartner identifies the top 10 business priorities for 2010. Retrieved 10 December 2010, from <http://www.gartner.com/it/page.jsp?id=1283413>, 2010.

Bibliography

- Michael Genrich, Alex Kokkonen, Jürgen Moormann, Michael zur Muehlen, Roger Tregear, Jan Mendling, and Barbara Weber. Challenges for business process intelligence: Discussions at the BPI workshop 2007. In A. H. M. ter Hofstede, Boualem Benatallah, and Hye-Young Paik, editors, *Proceedings of the 2nd International Workshop on Business Process Intelligence at BPM*, volume 4928 of *Lecture Notes in Computer Science*, pages 5–10, Berlin, Heidelberg, Germany, 2007. Springer.
- John Gerring. *Case study research: Principles and practices*. Cambridge University Press, Cambridge, UK, 2007.
- Stijn Goedertier. *Declarative techniques for modeling and mining business processes*. PhD thesis, Catholic University of Leuven, Belgium, 2008.
- M. Golani and S. Pinter. Generating a process model from a process audit log. In Wil M. P. van der Aalst, A. H. M. ter Hofstede, and M. Weske, editors, *Proceedings of the 3rd International Conference on Business Process Management*, pages 136–151, Berlin, Heidelberg, Germany, 2003. Springer.
- Hector Gonzalez, Jiawei Han, and Xiaolei Li. Mining compressed commodity workflows from massive RFID data sets. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pages 162–171, New York, NY, USA, 2006. ACM.
- Gianluigi Greco, Antonella Guzzo, and Luigi Pontieri. Discovering expressive process models by clustering log traces. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):1010–1027, 2006.
- Oliver Günther, Wolfhard Kletti, and Uwe Kubach. *RFID in manufacturing*. Springer, Berlin, Heidelberg, Germany, 2009.
- Michael Hammer and James Champy. *Reengineering the corporation: A manifesto for business revolution*. Harper Business, New York, NY, USA, 1993.
- Hans Robert Hansen. *Wirtschaftsinformatik I*. UTB, Stuttgart, Germany, 1996.
- H. Harrington. *Business process improvement. The breakthrough strategy for total quality, productivity, and competitiveness*. McGraw-Hill Professional, New York, NY, USA, 1991.
- Alexander Hars. *Referenzdatenmodelle - Grundlagen effizienter Datenmodellierung*. Gabler-Verlag, Wiesbaden, Germany, 1994.
- H. Heilmann. Workflow Management: Integration von Organisation und Informationsverarbeitung. In *Workflow Management*, number 176 in *Handbuch der Modernen Datenverarbeitung (HMD) - Praxis der Wirtschaftsinformatik*, pages 8–21, Heidelberg, Germany, 1994. dpunkt.verlag.

- J. C. Henderson and N. Venkatraman. Strategic alignment: leveraging information technology for transforming organizations. *IBM Systems Journal*, 38(2-3):472–484, 1999.
- K. W. Henning. *Einführung in die betriebswirtschaftliche Organisationslehre*. Springer, Berlin, Heidelberg, Germany, 1934.
- Joachim Herbst and Dimitris Karagiannis. Integrating machine learning and workflow management to support acquisition and adaptation of workflow models. In *Proceedings of the 9th IEEE International Workshop on Database and Expert Systems Applications*, pages 745–752, Washington, DC, USA, 1998. IEEE Computer Society. ISBN 0-8186-8353-8.
- S. Hume and D. Sunday. Fast string searching. *Software Practice and Experience*, 21(11):21–48, 1991.
- IDS Scheer AG. ARIS Process Performance Manager (ARIS PPM). IDS Scheer, Saarbrücken, Germany, Retrieved 01 March 2007, from <http://www.ids-scheer.com>, 2002.
- IDS Scheer AG. Retrieved 01 March 2007, from <http://www.ids-scheer.com>, 2007.
- IEEE Task Force on Process Mining. Standardization effort: event log format, retrieved 05 August 2010, from http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:standard_event_log_format, 2010.
- Jon Espen Ingvaldsen and Jon Atle Gulla. Preprocessing support for large scale process mining of SAP transactions. In *Proceedings of the International Workshop on Business Process Management*, pages 30–41, 2007.
- Iontas. Retrieved 01 March 2009, from <http://www.iontas.com>, 2009.
- Majid Iqbal and Michael Nieves. *Service strategy*. Stationery Office Books, London, UK, 2007.
- IT Governance Institute. Board briefing on IT governance. 2nd edition. Retrieved 12 August 2008, from www.itgi.org, 2003.
- IT Governance Institute. Control Objectives for Information and related Technology (COBIT), specification, version 4.1. Retrieved 12 April 2008, from <http://www.isaca.org>, 2007.
- IT Governance Institute. *IT Governance Global Status Report*. IT Governance Institute, Rolling Meadows, IL, USA, 2008a.
- IT Governance Institute. *IT Governance Global Status Report*, Retrieved 01 March 2009, from <http://www.isaca.de>. IT Governance Institute, Rolling Meadows, IL, USA, 2008b.

Bibliography

- IT Governance Institute. An executive view of IT governance. Retrieved 15 September 2009, from <http://www.isaca.org/Knowledge-Center/Research/Documents/An-Executive-View-of-IT-Gov-Research.pdf>, 2009.
- Lenka Ivantysynova. *RFID in manufacturing: Mapping the shop floor to IT-enabled business processes*. PhD thesis, Humboldt University of Berlin, Germany, 2008.
- Simon Jacobson, Jim Shepherd, Marianne D'Aquila, and Karen Carter. *The ERP market sizing report, 2006-2011*. AMR Research, Inc., Boston, MA, USA, 2007.
- John Jenkins, Paul Mills, Ludwig Loxley, and Bob Lilley. BRIDGE pharma traceability pilot. Preparation report. Retrieved 03 May 2009, from <http://www.bridge-project.eu/index.php/workpackage6/en>, 2007a.
- John Jenkins, Paul Mills, Robert Maidment, and Mike Profit. BRIDGE project. Business case report. Retrieved 03 December 2009, from http://www.bridge-project.eu/data/File/BRIDGE_WP06_Business_Case_Report.pdf, Mai 2007b.
- John Jenkins, Jeff Barnett, and Paul Mills. BRIDGE pharma traceability pilot. Evaluation report. Retrieved 03 November 2009, from <http://www.bridge-project.eu/index.php/workpackage6/en>, 2008.
- John Jenkins et al. BRIDGE pharma traceability pilot. Requirements analysis. Retrieved 13 December 2009, from <http://www.bridge-project.eu/index.php/public-deliverables/en/>, July 2007c.
- John Jenkins et al. BRIDGE project. Problem analysis. Retrieved 03 September 2009, from <http://www.bridge-project.eu/index.php/public-deliverables/en>, July 2007d.
- Wolfgang Johannsen and Matthias Goeken. IT-Governance - neue Aufgaben des IT-Managements. In *IT-Governance*, number 250 in Handbuch der Modernen Datenverarbeitung (HMD) - Praxis der Wirtschaftsinformatik, pages 7–20, Heidelberg, Germany, 2006. dpunkt.verlag.
- Wolfgang Johannsen and Matthias Goeken. *Referenzmodelle für IT Governance. Strategische Effektivität und Effizienz mit COBIT, ITIL & Co*. dpunkt.verlag, Heidelberg, Germany, 2007.
- Hank Johansson, Henry J. Johansson, and A. John Pendlebury. *Business process reengineering: Breakpoint strategies for market dominance*. John Wiley & Sons, Inc., New York, NY, USA, 1993.
- Yeh-Chun Juan. A string comparison approach to process logic differences between business process models. In *Proceedings of the 9th Joint Conference on Information Sciences*, pages 909–912, Amsterdam, The Netherlands, 2006. Atlantis Press.

- Henning Kagermann and Hubert Österle. *Geschäftsmodelle 2010, Wie CEOs Unternehmen transformieren*. Frankfurter Allgemeine Zeitung, Frankfurt a. M., Germany, 2006.
- R. Kaplan and D. Norton. The balanced scorecard - measures that drive performance. *Harvard Business Review*, 70(1):71–79, 1992.
- Ramisa Kashanchi and Janet Toland. Can ITIL contribute to IT/business alignment? an initial investigation. *Wirtschaftsinformatik*, 48(5):340–348, 2006.
- Gamal Kassem. *Application Usage Mining: Grundlagen und Verfahren*. PhD thesis, Otto-von-Guericke-University, Magdeburg, Germany, 2007.
- G. Keller and T. Teufel. *SAP R/3 process oriented implementation*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1998a.
- G. Keller and T. Teufel. *SAP R/3 process oriented implementation: Iterative process prototyping*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1998b.
- Allen Kent and James G. Williams. *Encyclopedia of computer science and technology*, volume 40. Marcel Dekker, Inc., New York, NY, USA, 1999.
- William J. Kettinger and James T. C. Teng. Business process change: A study of methodologies, techniques, and tools. *MIS Quarterly*, 21(1):55–80, March 1997.
- Rajiv Kohli and Ellen Hoadley. Towards developing a framework for measuring organizational impact of IT-enabled BPR: case studies of three firms. *SIGMIS Database*, 37:40–58, 2006.
- Jochen Malte Küster, Jana Koehler, and Ksenia Ryndina. Improving business process models with reference models in business-driven development. In *Proceedings of the 2nd Business Process Design Workshop at BPM*, volume 4103 of *Lecture Notes in Computer Science*, pages 35–44, Berlin, Heidelberg, Germany, 2006. Springer.
- Martin Kütz. *Kennzahlen in der IT, Werkzeuge für das Controlling und Management*. dpunkt.verlag, Heidelberg, Germany, 2009.
- Martin Lang. *Prozess-Mining und Prozessbewertung zur Verbesserung klinischer Workflows im Umfeld bildergezeugender Fächer*. PhD thesis, Erlangen-Nürnberg University of Technology, Erlangen-Nürnberg, Germany, 2008.
- Hau L. Lee, V. Padmanabhan, and Seungjin Whang. The bullwhip effect in supply chains. *Sloan Management Review*, 38(3):93–102, 1997.
- Alexis Leon. *Enterprise resource planning*. Tata McGraw-Hill Publishing Company Limited, New Delhi, IND, 2007.
- I. Levenshtein. Binary code capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, 10(8):707–710, 1966.

Bibliography

- F. Leymann and W. Altenhuber. Managing business processes as an information resource. *IBM Systems Journal*, 33(2):326–348, 1994.
- Robert Lilley, Phil Parkins, Tony Walsh, and John Jenkins. BRIDGE project. Deployment report. Retrieved 03 November 2009, from http://www.bridge-project.eu/data/File/BRIDGE_WP06_Pilot_Deployment_Report.pdf, July 2008.
- D. Lin, H. G. Elmongui, E. Bertino, and B. C. Ooi. Data management in RFID applications. In *Proceedings of the 18th International Conference on Database and Expert Systems Applications*, volume 4653 of *Lecture Notes in Computer Science*, pages 434–444, Berlin, Heidelberg, Germany, 2007. Springer.
- G. W. Loveman. Assessing the productivity impact of information technologies. In T. J. Allen and M. Scott-Morton, editors, *Information Technology and the Corporation of the 1990s*, pages 84–110, New York, NY, USA, 1994. Oxford University Press.
- D. Luckham. *The power of events: An introduction to complex event processing in distributed enterprise systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
- Brian E. Mansir and Nicholas R. Schacht. *An introduction to the continuous improvement process: Principles and practices*. Logistics Management Institute, Bethesda, MD, USA, 1989.
- Joachim Melcher and Detlef Seese. Process measurement: Insights from software measurement on measuring process complexity, quality, and performance. Technical report, Karlsruhe University of Technology, Karlsruhe, Germany, 2008.
- J. Mendling, H. M. W. Verbeek, B. F. van Dongen, Wil M. P. van der Aalst, and G. Neumann. Detection and prediction of errors in EPCs of the SAP reference model. *Data and Knowledge Engineering*, 64(1):312–329, 2008. ISSN 0169-023X.
- Jan Mendling. Testing density as a complexity metric for EPCs. Technical report, German EPC workshop on density of process models, 2006.
- Jan Mendling and Carlo Simon. Business process design by view integration. In *Proceedings of the 2nd Workshop on Business Processes Design at BPM*, volume 4103 of *Lecture Notes in Computer Science*, pages 55–64, Berlin, Heidelberg, Germany, 2006. Springer.
- Jan Mendling, Gustaf Neumann, and Markus Nüttgens. Yet another event-driven process chain. In *Proceedings of the 3rd International Conference on Business Process Management*, volume 3649 of *Lecture Notes in Computer Science*, pages 428–433, Berlin, Heidelberg, Germany, 2005. Springer.
- Christof Menzies. *Sarbanes-Oxley Act. Professionelles Management interner Kontrollen*. Poeschel Verlag, Stuttgart, Germany, 2004.

- Matthias Meyer, Rüdiger Zarnekow, and Lutz Kolbe. IT-Governance – Begriff, Status quo und Bedeutung. *Wirtschaftsinformatik*, pages 445–448, 2003.
- Jans Mieke, Nadine Lybaert, and Koen Vanhoof. Business process mining for internal fraud risk reduction: Results of a case study. In *Proceedings of the 2nd Workshop on Induction of Process Models at ECML*, 2008.
- Robin Milner. An algebraic definition of simulation between programs. Technical report, Stanford University, Stanford, CA, USA, 1971.
- B.S. Morgan and W.A. Schiemann. Measuring people and performance: Closing the gaps. *Quality Progress*, 32(1):47–53, 1999.
- H.R. Motahari-Nezhad, R. Saint-Paul, B. Benatallah, and F. Casati. Protocol discovery from imperfect service interaction logs. In *Proceedings of the 23rd International Conference on Data Engineering*, pages 1405–1409, 2007.
- Laura Mărușter and Nick van Beest. Redesigning business processes: A methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, 21(3):267–297, 2009.
- Laura Mărușter, J. C. Wortmann, A. J. M. M. Weijters, and Wil M. P. van der Aalst. Discovering distributed processes in supply chains. In Jagdev, H., Wortmann, H., Pels, H., and Hirnschall, A., editors, *Proceedings of the 8th International Conference on Advanced Production Management Systems*, pages 119–128, 2002.
- T. Murata. Petri nets: Properties, analysis, and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989.
- Gonzalo Navarro. A guided tour to approximate string matching. *ACM Transactions on Computer Survey*, 33(1):31–88, 2001. ISSN 0360-0300.
- Saul Ben Needleman and Christian Dennis Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, March 1970.
- Stefan Neumann, Christian Probst, and Clemens Wernsmann. *Process management: A guide for the design of business processes*, chapter Continuous process management, pages 233–250. Springer, Berlin, Heidelberg, Germany, 2003.
- Fred Niederman, Richard G. Mathieu, Roger Morley, and Ik-Whan Kwon. Examining RFID applications in supply chain management. *ACM Transactions on Software Engineering and Methodology*, 50(7):92–101, 2007.
- F. Nordsieck. Grundprobleme und Grundprinzipien der Organisation des Betriebsaufbaus. *Die Betriebswirtschaft*, 24(6):158–162, 1931.
- Fritz Nordsieck. *Grundlagen der Organisationslehre*. Poeschel Verlag, Stuttgart, Germany, 1934.

Bibliography

- OASIS. Web Services Business Process Execution Language (WSBPEL), specification, version 2.0. Retrieved 02 February 2009, from <http://www.oasis-open.org>, January 2007.
- Object Management Group. Business Process Modeling Notation (OMG BPMN), specification, version 1.1. Retrieved 12 September 2009, from <http://www.omg.org/spec/BPMN/1.1/>, February 2006.
- Object Management Group. Business Process Maturity Model (OMG BPMM), specification, version 1.0. Retrieved 11 December 2008, from <http://www.omg.org/spec/BPMM>, 2008.
- Office of Government Commerce. *The official introduction to the ITIL service lifecycle*. Stationery Office Books, London, UK, 2007.
- OpenConnect. Retrieved 03 November 2009, from <http://www.oc.com>, 2009.
- Hubert Österle. *Business in the information age - heading for new processes*. Springer, Berlin, Heidelberg, Germany, 1995.
- Pallas Athena. Retrieved 03 December 2010, from <http://www.pallas-athena.com>, 2010.
- W.D. Pauw, R. Hoch, and Huang. Discovering conversations in web services using semantic correlation analysis. In *Proceedings of the 5th IEEE International Conference on Web Services*, pages 639–646, Washington, DC, USA, 2007. IEEE Computer Society.
- M. Pesic and Wil M. P. van der Aalst. Towards a reference model for work distribution in workflow management systems. In *Proceedings of the 3rd International Workshop on Business Process Reference Models at BPM*, pages 30–44, 2005.
- J. L. Peterson. *Petri net theory and the modeling of systems*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1981.
- R. Peterson. Information strategies and tactics for information technology governance. In *Strategies for Information Technology Governance*, Hershey, PA, USA, 2004. Idea Group.
- C.A. Petri. *Kommunikation mit Automaten*. PhD thesis, Bonn University, Bonn, Germany, 1962.
- K. Pirklbauer. A study of pattern-matching algorithms. *Structured Programming*, 13(2):89–98, 1992.
- Amir Pnueli. The temporal logic of programs. In *Proceedings of the 18th IEEE Annual Symposium on Foundations of Computer Science*, pages 46–57, Washington, DC, USA, 1977. IEEE Computer Society.

- Lucia Pomello, Grzegorz Rozenberg, and Carla Simone. A survey of equivalence notions for net based systems. In *Proceedings of the 11th International Advances in Petri Nets 1992*, volume 609 of *Lecture Notes in Computer Science*, pages 410–472, Berlin, Heidelberg, Germany, 1992. Springer.
- M. E. Porter and V. E. Millar. How information gives you competitive advantage. *Harvard Business Review*, 63(4):149–160, 1985.
- Michael E. Porter. *Competitive advantage. Creating and sustaining superior performance*. The Free Press, New York, NY, USA, 1985.
- Dorian Pyle. *Data preparation for data mining*. Morgan Kaufmann Publishers, Inc., San Francisco, CA, USA, 1999.
- Abirami Radhakrishnana, Xingxing Zub, and Varun Groverc. A process-oriented perspective on differential business value creation by information technology: An empirical investigation. *Omega*, 36(6), 2008.
- Michael Rasmussen. *Overcoming risk and compliance myopia*. Forrester Research, Inc., Cambridge, MA, USA, 2006.
- T. Ravichandran and C. Lertwongsatien. Effect of information systems resources and capabilities on firm performance: A resource-based perspective. *Journal of Management Information Systems*, 21:237–276, 2005.
- Marcel Redinger. Implement an ITIL employee service desk with IC WebClient . *SAP Informatik-Spektrum*, 3:30–39, 2007.
- G. Regev, P. Soffer, and R. Schmidt. Taxonomy of flexibility in business processes. In *Proceedings of the 7th Workshop on Development and Support at BPMDS*, pages 90–93, 2006.
- Hajo A. Reijers. *Process-aware information systems: Bridging people and software through process technology*, chapter Process design and redesign, pages 207–234. John Wiley & Sons, Inc., New York, NY, USA, 2005.
- Hajo A. Reijers and Irene Vanderfeesten. Cohesion and coupling metrics for workflow process design. In *Proceedings of the 2nd International Conference on Business Process Management*, volume 3080 of *Lecture Notes in Computer Science*, pages 290–305, Berlin, Heidelberg, Germany, 2004. Springer.
- Wolfgang Reisig and Grzegorz Rozenberg. Lectures on Petri nets I: Basic models, advances in Petri nets. In *Petri Nets (2)*, volume 1491 of *Lecture Notes in Computer Science*, pages 529–586, London, UK, 1998. Springer. ISBN 3-540-65306-6.
- R. Rolles. *Verbesserung von Geschäftsprozessen mit flexiblen Workflow-Management-Systemen*, chapter Kontinuierliche Verbesserung von workflow gestützten Geschäftsprozessen, pages 109–133. Physica-Verlag, Heidelberg, Germany, 1998.

Bibliography

- M. Rosemann and Wil M. P. van der Aalst. A configurable reference modelling language. *Information Systems*, 32(1):1–23, 2007. ISSN 0306-4379.
- A. Rozinat and W. M. P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008. ISSN 0306-4379.
- A. Rozinat, Ana Karla Alves de Medeiros, C.W. Günther, A. J. M. M. Weijters, and Wil M. P. van der Aalst. Towards an evaluation framework for process mining algorithms. Technical Report WP 224, BETA Working Paper Series, Eindhoven University of Technology, Eindhoven, The Netherlands, 2007a.
- A. Rozinat, I.S.M. de Jong, C.W. Günther, and Wil M. P. van der Aalst. Process mining of test processes: A case study. Technical Report WP 220, BETA Working Paper Series, Eindhoven University of Technology, Eindhoven, The Netherlands, 2007b.
- A. Rozinat, M. Veloso, and Wil M. P. van der Aalst. Evaluating the quality of discovered process models. In W. Bridewell, T. Calders, et al., editors, *Proceedings of the 2nd International Workshop on the Induction of Process Models at ECML*, pages 45–52, 2008.
- A. Rozinat, I.S.M. de Jong, C.W. Günther, and Wil M. P. van der Aalst. Process mining applied to the test process of wafer steppers in ASML. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 39(4):474–479, 2009.
- Anne Rozinat and Wil M. P. van der Aalst. Conformance testing: Measuring the fit and appropriateness of event logs and process models. In Christoph Bussler and Armin Haller, editors, *Proceedings of the 2nd International Workshop on Business Process Intelligence at BPM*, volume 3812 of *Lecture Notes in Computer Science*, pages 163–176, Berlin, Heidelberg, Germany, 2006. Springer.
- N. Russell, A. H. M. ter Hofstede, Wil M. P. van der Aalst, and N. Mulyar. Workflow control-flow patterns: A revised view. Technical report, BPM Center, 2006.
- Stefan Sackmann and Martin Kähler. ExPDT: A layer-based approach for automating compliance. *Wirtschaftsinformatik*, 50(5):366–374, 2008.
- SAP AG. Retrieved 03 November 2008, from <http://www.sap.com>, 2008.
- August-Wilhelm Scheer. *Enterprise-wide data modelling: Information systems in industry*. Springer, Berlin, Heidelberg, Germany, 1989.
- August-Wilhelm Scheer. *Business process engineering: Reference models for industrial enterprises*. Springer, Berlin, Heidelberg, Germany, 1994.
- August-Wilhelm Scheer. *Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse*. Springer, Berlin, Heidelberg, Germany, 1997.

- August-Wilhelm Scheer, W. Hoffmann, and R. Wein. Customizing von Standardsoftware mit Referenzmodellen. In *Objektorientierte Anwendungsarchitekturen - unternehmensweit*, number 31 in Handbuch der Modernen Datenverarbeitung (HMD) - Praxis der Wirtschaftsinformatik, pages 92–103, Heidelberg, Germany, 1994. dpunkt.verlag.
- Gerhard Schewe. *Unternehmensverfassung: Corporate Governance im Spannungsfeld von Leitung, Kontrolle und Interessenvertretung*. Springer, Berlin, Heidelberg, Germany, 2005.
- Igor E.A. Segers. Deloitte enterprise risk services: Investigating the application of process mining for auditing purpose. Master’s thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2007.
- Aditya Sharma, Alex Citurs, and Benn Konsynski. Strategic and institutional perspectives in the adoption and early integration of radio frequency identification (RFID). In *Proceedings of the 40th IEEE International Conference on System Sciences*, pages 224–233, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 0-7695-2755-8.
- Asuman Suenbuel and Ming-Shien Shan. Towards enterprise archeology: Extracting business processes from runtime event data. In *Proceedings of the 6th IEEE International Conference on Data Mining*, pages 468–473, Washington, DC, USA, 2006. IEEE Computer Society.
- Supply Chain Operations Council. Supply Chain Operations Reference Model (SCOR), specification, version 8.0. Retrieved 10 May 2007, from <http://www.supply-chain.org>, 2006.
- Paul P. Tallon, Kenneth L. Kraemer, and Vijay Gurbaxani. Executives’ perceptions of the business value of information technology: a process-oriented approach. *Journal of Management Information Systems*, 16(4):145–173, 2000. ISSN 0742-1222.
- Gerrit Tamm and Christoph Tribowski. *RFID (Informatik im Fokus)*. Springer, Berlin, Heidelberg, Germany, 2010.
- Frederick Winslow Taylor. *The principles of scientific management*. Harper and Brothers, New York, NY, USA, 1911.
- Sharon Taylor, David Cannon, and David Wheeldon. *Service operation*. Stationery Office Books, London, UK, 2007a.
- Sharon Taylor, Gary Case, and George Spalding. *Continual service improvement*. Stationery Office Books, London, UK, 2007b.
- Sharon Taylor, Shirley Lacy, and Ivor Macfarlane. *Service transition*. Stationery Office Books, London, UK, 2007c.
- Sharon Taylor, Colin Rudd, and Vernon Lloyd. *Service design*. Stationery Office Books, London, UK, 2007d.

Bibliography

- Rolf Alexander Teubner. *Organisations- und Informationssystemgestaltung*. PhD thesis, Münster University, Münster, Germany, 1999.
- The BRIDGE Project. Retrieved 03 May 2009, from <http://www.bridge-project.eu>, 2009.
- Oliver Thomas. Understanding the term reference model in information systems research: History, literature analysis, and explanation. In *Proceedings of the 1st International Workshop on Business Process Reference Models at BPM*, volume 3812 of *Lecture Notes in Computer Science*, pages 484–496, Berlin, Heidelberg, Germany, 2006. Springer.
- J. van Bon. *IT Service Management: An Introduction*. Van Haren Publishing, Zaltbommel, The Netherlands, 2002.
- Jan van Bon and Tienieke Verheijen. *Frameworks für das IT Management*. Van Haren Publishing, Zaltbommel, The Netherlands, 2007.
- Jan van Bon, Arjen de Jong, Mike Piper, Ruby Tjassing, Tienieke Verheijen, and Annelies van der Veen, editors. *IT service management global best practices*. Van Haren Publishing, Zaltbommel, The Netherlands, 2008.
- W. M. P. van der Aalst, H. T. de Beer, and B. F. van Dongen. Process mining and verification of properties: An approach based on temporal logic. In Robert Meersman and Zahir Tari, editors, *Proceedings of the 7th International Conference on Cooperative Information Systems at OTM*, volume 3760 of *Lecture Notes in Computer Science*, pages 130–147, Berlin, Heidelberg, Germany, 2005. Springer.
- Wil M. P. van der Aalst. Verification of workflow nets. In *Proceedings of the 18th International Conference on Application and Theory of Petri Nets*, volume 1248 of *Lecture Notes in Computer Science*, pages 407–426, London, UK, 1997. Springer. ISBN 3-540-63139-9.
- Wil M. P. van der Aalst. Formalization and verification of event-driven process chains. *Information and Software Technology*, 41(10):639–650, 1999.
- Wil M. P. van der Aalst. Inheritance of business processes: A journey visiting four notorious problems. In H. Ehrig, W. Reisig, et al., editors, *Proceedings of the 21st International Conference on Applied Informatics on Petri Net Technology for Communication-Based Systems*, volume 2472 of *Lecture Notes in Computer Science*, pages 383–408, Berlin, Heidelberg, Germany, 2003. Springer.
- Wil M. P. van der Aalst. Business alignment: Using process mining as a tool for delta analysis and conformance testing. *Requirements Engineering*, 10(3):198–211, 2005. ISSN 0947-3602.
- Wil M. P. van der Aalst and Ana Karla Alves de Medeiros. Process mining and security: Detecting anomalous process executions and checking process conformance. In *Electronic Notes in Theoretical Computer Science*, pages 3–21, 2005.

- Wil M. P. van der Aalst and Twan Basten. Identifying commonalities and differences in object life cycles using behavioral inheritance. In *Proceedings of the 22nd International Conference on Application and Theory of Petri Nets*, volume 2075 of *Lecture Notes in Computer Science*, pages 32–52, London, UK, 2001. Springer. ISBN 3-540-42252-8.
- Wil M. P. van der Aalst and Boudewijn F. van Dongen. Discovering workflow performance models from timed logs. In *Proceedings of the 1st International Conference on Engineering and Deployment of Cooperative Information Systems*, pages 45–63, London, UK, 2002. Springer. ISBN 3-540-44222-7.
- Wil M. P. van der Aalst and A. J. M. M. Weijters. Process mining: A research agenda. *Computers in Industry*, 53(3):231–244, 2004. ISSN 0166-3615.
- Wil M. P. van der Aalst and A. J. M. M. Weijters. *Process-aware information systems: Bridging people and software through process technology*, chapter Process mining, pages 235–256. John Wiley & Sons, Inc., New York, NY, USA, 2005.
- Wil M. P. van der Aalst, Arthur H. M. ter Hofstede, and Mathias Weske. Business process management: A survey. In *Proceedings of the 3rd International Conference on Business Process Management*, volume 2678 of *Lecture Notes in Computer Science*, pages 1–12, Berlin, Heidelberg, Germany, 2003a. Springer.
- Wil M. P. van der Aalst, Boudewijn F. van Dongen, J. Herbst, L. Mărușter, G. Schimm, and A. J. M. M. Weijters. Workflow mining: A survey of issues and approaches. *Data and Knowledge Engineering*, 47(2):237–267, 2003b.
- Wil M. P. van der Aalst, A. J. M. M. Weijters, and Laura Mărușter. Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004.
- Wil M. P. van der Aalst, A. Dreiling, F. Gottschalk, M. Rosemann, and M. Jansen-Vullers. Configurable process models as a basis for reference modeling. In C. Bussler and A. Haller, editors, *Proceedings of the 2nd International Workshop on Business Process Intelligence at BPM*, volume 3812 of *Lecture Notes in Computer Science*, pages 512–518, Berlin, Heidelberg, Germany, 2006. Springer.
- Wil M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, Boudewijn F. van Dongen, Ana Karla Alves de Medeiros, M. Song, and H. M. W. Verbeek. Business process mining: An industrial application. *Information Systems*, 32(5):713–732, 2007a.
- Wil M. P. van der Aalst, Boudewijn F. van Dongen, C. W. Günther, R. S. Mans, Ana Karla Alves de Medeiros, A. Rozinat, V. Rubin, M. Song, H. M. W. Verbeek, and A. J. M. M. Weijters. ProM 4.0: Comprehensive support for real process analysis. In *Proceedings of the 28th International Conference on Application and Theory of Petri Nets and other Models of Concurrency*, volume 4546 of *Lecture Notes in Computer Science*, pages 484–494, Berlin, Heidelberg, Germany, 2007b. Springer.

Bibliography

- Wil M. P. van der Aalst, M. Dumas, Chun Ouyang, Anne Rozinat, and Eric Verbeek. Conformance checking of service behavior. *ACM Transactions on Internet Technology*, 8(3):1–30, 2008. ISSN 1533-5399.
- B. F. van Dongen, J. Mendling, and W. M. P. van der Aalst. Structural patterns for soundness of business process models. In *Proceedings of the 10th IEEE International Enterprise Distributed Object Computing Conference*, pages 116–128, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2558-X.
- Boudewijn F. van Dongen and Wil M. P. van der Aalst. A meta model for process mining data. In Oscar Pastor, editor, *Proceedings of the 17th International Conference on Advanced Information Systems Engineering*, volume 3520 of *Lecture Notes in Computer Science*, pages 309–320, 2005.
- Boudewijn F. van Dongen, Ana Karla Alves de Medeiros, H.M.W. Verbeek, A. J. M. M. Weijters, and Wil M. P. van der Aalst. The ProM framework: A new era in process mining tool support. In Gianfranco Ciardo and Philippe Darondeau, editors, *Proceedings of the 26th International Conference of Applications and Theory of Petri Nets*, volume 3536 of *Lecture Notes in Computer Science*, pages 444–454, Berlin, Heidelberg, Germany, 2005. Springer.
- Boudewijn F. van Dongen, Remco Dijkman, and Jan Mendling. Measuring similarity between business process models. In B. Thalheim, editor, *Proceedings of the 20th International Conference on Advanced Information Systems Engineering*, volume 5074 of *Lecture Notes in Computer Science*, pages 450–464, Berlin, Heidelberg, Germany, 2008. Springer.
- M. van Giessel. Process mining in SAP R/3: A method for applying process mining to SAP R/3. Master’s thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2004.
- Rob J. van Glabbeek and W. Peter Weijland. Branching time and abstraction in bisimulation semantics. *Communications of the ACM*, 43(3):555–600, 1996. ISSN 0004-5411.
- J. van Reenen and R. Sadun. Information technology and productivity: It ain’t what you do it’s the way that you do I.T. *EDS Innovation Research Programme, London School of Economics. Discussion Paper Series*, 2005.
- H. M. W. Verbeek, Boudewijn F. van Dongen, J. Mendling, and Wil M. P. van der Aalst. Interoperability in the ProM framework. In *Proceedings of the International Workshop and Doctoral Consortium at CAiSE*, pages 619–630, 2006.
- H.M.W. Verbeek. *Verification and enactment of workflow management systems*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2004.
- Jan vom Brocke. *Referenzmodellierung : Gestaltung und Verteilung von Konstruktionsprozessen*. Logos, Berlin, Germany, 2003.

- T. F. Wallace and M. H. Kremzar. *ERP: Making it happen: The implementers' guide to success with enterprise resource planning*. John Wiley & Sons, Inc., New York, NY, USA, 2001.
- Fusheng Wang and Peiya Liu. Temporal management of RFID data. In *Proceedings of the 31st International Conference on Very large Data Bases*, pages 1128–1139, Danvers, MA, USA, 2005. VLDB Endowment, Inc. ISBN 1-59593-154-6.
- A. J. M. M. Weijters, Wil M. P. van der Aalst, and Ana Karla Alves de Medeiros. Process mining with the Heuristics miner algorithm. Technical Report WP 166, BETA Working Paper Series, Eindhoven University of Technology, Eindhoven, The Netherlands, 2006.
- Peter Weill and Jeanne W. Ross. *IT governance: How top performers manage IT decision rights for superior results*. Harvard Business School Press, Boston, MA, USA, 2004.
- Mathias Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer, Berlin, Heidelberg, Germany, 2007.
- T. Wilde and T. Hess. Forschungsmethoden der Wirtschaftsinformatik - eine empirische Untersuchung. *Wirtschaftsinformatik*, 4:280–287, 2007.
- Robert Winter. *Business Engineering: Auf dem Weg zum Unternehmen des Informationszeitalters*, chapter Modelle, Techniken und Werkzeuge im Business Engineering. Springer, Berlin, Heidelberg, Germany, 2003.
- Michael zur Muehlen. Process-driven management information systems - combining data warehouses and workflow technology. In *Proceedings of the 4th International Conference on Electronic Commerce Research*, pages 550–566, Dallas, TX, USA, 2001.
- Michael zur Muehlen. *Workflow-based process controlling. Foundation, design, and application of workflow-driven process information systems*. Logos Verlag, Berlin, Germany, 2004.
- Michael zur Muehlen and M. Rosemann. Workflow-based process monitoring and controlling - technical and organizational issues. In R. Sprague, editor, *Proceedings of the 33rd IEEE Hawaii International Conference on System Science*, pages 1–10, Los Alamitos, CA, USA, 2000. IEEE Computer Society.

Author's Publications

- Kerstin Gerke. Impact of data aggregation on the significance of process mining results: An experimental evaluation. In *Proceedings of the 3rd IEEE Workshop on Agile Cooperative Process-Aware Information Systems at WETICE*, pages 247–248, Washington, DC, USA, 2008. IEEE Computer Society.
- Kerstin Gerke and Gerrit Tamm. Process mining and data aggregation strategies for business collaboration networks. *Communications of Arab Computer Society (ACS)*, 1(1):10–19, 2008.
- Kerstin Gerke and Gerrit Tamm. Qualitätsmanagement zur Steuerung von IT-Prozessen auf der Basis von Referenzmodellen und Process-Mining. In Stefan Reinheimer, editor, *Prozessmanagement*, number 266 in Handbuch der Modernen Datenverarbeitung (HMD) - Praxis der Wirtschaftsinformatik, pages 43–51, Heidelberg, Germany, 2009a. dpunkt.verlag.
- Kerstin Gerke and Gerrit Tamm. Continuous quality improvement of IT processes based on reference models and process mining. In *Proceedings of the 15th Americas Conference on Information Systems*. AIS Electronic Library, Paper 786, 2009b.
- Kerstin Gerke, Jorge Cardoso, and Alexander Claus. Measuring the compliance of processes with reference models. In R. Meersman, T. Dillon, and P. Herrero, editors, *Proceedings of the 17th International Conference on Cooperative Information Systems at OTM*, volume 5870 of *Lecture Notes in Computer Science*, pages 76–93, Berlin, Heidelberg, Germany, 2009a. Springer.
- Kerstin Gerke, Jorge Cardoso, and Alexander Claus. *System and Method of Measuring Process Compliance*. U.S. Patent No. 000005-013700. Washington, D.C, USA: U.S. Patent and Trademark Office, 2009b.
- Kerstin Gerke, Jan Mendling, and Alexander Claus. Process mining of RFID-based supply chains. In *Proceedings of the 11th IEEE International Conference on Commerce and Enterprise Computing*, pages 285–292, Washington, DC, USA, 2009c. IEEE Computer Society.
- Kerstin Gerke, Jan Mendling, and Konstantin Tarmyshov. Case construction for mining supply chain processes. In Witold Abramowicz, editor, *Proceedings of the 12th International Conference on Business Information Systems*, volume 21 of *Lecture Notes in Business Information Processing*, pages 181–192, Berlin, Heidelberg, Germany, 2009d. Springer.

Author's Publications

Kerstin Gerke, Konstantin Petruch, and Gerrit Tamm. *Quality management for IT services: Perspectives on business and process performance*, chapter Continual process improvement based on ITIL and process mining, pages 144–166. IGI Global Business Science Reference, Hershey, PA, USA, 2010a.

Kerstin Gerke, Konstantin Petruch, and Gerrit Tamm. Optimization of service delivery through continual process improvement: A case study. In *Proceedings of the 2nd International Symposium on Services Science*, Lecture Notes in Informatics from the Germany Computer Science Society, Berlin, Heidelberg, Germany, 2010b. Springer.

List of Figures

2.1. Life cycle of business process management	16
2.2. Some examples of modeling notations	21
2.3. Strategic alignment model according to ITIL	37
2.4. Differentiation between IT management, ITSM, and IT governance	40
2.5. Examples of RFID tags	41
2.6. Examples of RFID readers	42
2.7. Example of an RFID architecture	43
2.8. URN notation of the Uniform Resource Identifier	43
2.9. EPCIS events and fields	44
3.1. Mining Extensible Markup Language format	50
3.2. Focus shifts of EPCIS events	53
3.3. Composition within an business object	65
3.4. Associations among business objects	68
3.5. Implementation of algorithm EPCIS2MXML	79
4.1. Automobile supply chain	88
4.2. Pharmaceutical supply chain	89
4.3. Application Interaction Center	92
4.4. Complaint handling process of a passenger airline	93
4.5. Validation of algorithm EPCIS2MXML	97
4.6. Event log opened in ProM	98
4.7. Process model Aut-1 of the overall product flow	99
4.8. Process model Aut-1 of completed instances	100
4.9. Excerpt of models M'_3 and M_2 of Aut-1	101
4.10. Runtime of algorithm EPCIS2MXML	103
4.11. Process model Aut-1 of the first tier supplier	104
4.12. Process model Aut-1 with TransactionEvents	105
4.13. Process model Aut-1 of loudspeakers	106
4.14. Process model Aut-1 of returnable items and doors	107
4.15. Process model Pha-1 of the pharmaceutical supply chain	109
4.16. Process model Pha-1 of manufacturer Athlone	113
4.17. Log Inspector plug-in of model Air-1 in ProM	115
4.18. Process model Air-1 of the complaint handling	116
4.19. Process model Air-1 of group Cust. Payments	118
4.20. Process model Air-1 on transaction level	119
4.21. Process model Tel-1 of service e-mail on action level	121

List of Figures

4.22. Service operation of an ISP	122
4.23. Differences between actions and activities	123
4.24. Process model Tel-1 of service e-mail on activity level	125
4.25. Process models Tel-1 on different granularity levels	126
4.26. Data preparation effort	127
4.27. Data preparation effort and process model's quality	129
5.1. Procedure model for service operation through process mining	135
5.2. IT's contribution to efficiency, innovation, and strategy	138
5.3. Maturity of IT management	138
5.4. Challenges in IT management	139
5.5. Outcomes of IT management	139
5.6. As-is process modeling	140
5.7. Process performance analysis	140
5.8. Awareness of reference models	141
5.9. Expected benefits from reference models	142
5.10. Usefulness of process mining	142
5.11. Entities of a compliance analysis	143
5.12. Continual process improvement approach	144
5.13. Extended CPSI approach	146
5.14. Maturity assessment of the incident management	147
5.15. Model hierarchy Air-1 for complaint management	149
5.16. Conformance Checker plug-in of model Air-1	150
5.17. To-be process model Tel-1	152
5.18. Process models Tel-2 and Tel-3	153
5.19. Key performance indicator selection process	154
5.20. Conformance Checker plug-in of model Tel-1	156
5.21. Deviations between log and process model Tel-1	157
5.22. Conformance Checker plug-in of model Tel-1 showing the result after the adaptation was done	157
6.1. Receipt of complaints at an airline	168
6.2. Customization of a reference model	170
6.3. Average runtime	180
6.4. Runtime as function of activities	180
6.5. Two process models used in the running example	181
6.6. Process model Air-2 used for compliance analysis	186
6.7. Sequence-Based Compliance Analysis plug-in	187
6.8. Structural Precision/Recall plug-in Air-1	188
6.9. Footprint Similarity plug-in Air-1	189
6.10. Sequenced-based Compliance Analysis of Tel-1	190
6.11. Footprint Similarity plug-in Tel-1	191
6.12. Sequenced-based Compliance Analysis in Pha-1	192
6.13. Sequenced-based Compliance Analysis in Pha-1 - details	192

6.14. Footprint Similarity plug-in Pha-1	193
B.1. Responsibility for IT management	213
B.2. Resources committed to process improvement and process communication	213
B.3. Subjects of measurement	214
B.4. Behavior of the IT department vis-a-vis IT	214
B.5. Behavior of the IT department vis-a-vis BPM	215
B.6. Means of performance evaluation	216
B.7. Effects of IT improvements on business	216
B.8. Awareness of reference models	217
B.9. Use of reference models	217
B.10. Purpose of reference models' use	218
B.11. Characteristics of reference models	218
B.12. Means for maintaining reference models	219
B.13. Experiences with reference models' use	220
B.14. Awareness of process mining	220
B.15. Helpfulness of automatic construction of process models	220
B.16. Awareness of process mining's products	221
B.17. Barriers to industrial application of process mining	221
C.1. Process model Tel-1 of the transaction scheme	223
C.2. Process model Tel-2 of service S_1 on action level	224
C.3. Process model Tel-2 of service S_1 on activity level	225
C.4. Process model Tel-3 of service S_2 on action level	226
C.5. Process model Tel-3 of service S_2 on activity level	227
C.6. Process model Tel-4 of service S_3 on action level	228
C.7. Process model Tel-4 of service S_3 on activity level	229
D.1. Example Heuristics net	236
E.1. SCOR process category D2	237

List of Tables

2.1. Categorization of reference models	26
3.1. Example MXML event derived from RFID events	58
3.2. Mapping of event fields	59
3.3. Activities extracted from a CRM activity	70
3.4. Example MXML events derived from enterprise data	70
3.5. Document flow in the SAP CRM system	73
4.1. Organizations of use cases	85
4.2. Properties of use cases	86
4.3. Behavior of nodes	87
4.4. Events generated based on behavior	87
4.5. Fragment of the original history of action	94
4.6. Example MXML event derived from the history of action	120
5.1. Excerpt of the indicator-based measurement system	155
5.2. Indicator candidates for inefficiency	155
6.1. Overview of current verification measures in the context of process mining	172

Acronyms

AML	ARIS Markup Language
AR	Adherence to references
ARIS	Architecture of Integrated Information Systems
ARIS PPM	ARIS Process Performance Manager
ASL	Application Services Library
ASN	Advanced shipping notification
Athlone	Athlone Laboratories
Avon	Avonmouth
BAPI	Business Application Programming Interface
Barts	Bart and The London NHS Trust
BDoc	Business document
BiSL	Business Information Service Library
BO	Business object
BOL	Business object layer
BOR	Business object repository
BPE	Business process engineering
BPEL	Business process execution language
BPM	Business process management
BPMI	Business Process Management Initiative
BPMM	Business process maturity model
BPMN	Business process modeling notation
BPR	Business process reengineering
BPRE	Business process reverse engineering
BRIDGE	Building Radio Frequency Identification for the Global Environment
COBIT	Control Objectives for Information and related Technology
Camb	Cambridge
CCTA	Central Computer and Telecommunications Agency
Cd	Compliance degree
CIO	Chief information officer
Cm	Compliance maturity
CMMI	Capability Maturity Model Integration

Acronyms

COSO	The Committee of Sponsoring Organizations of the Treadway Commission Internal Control
CPG	CPG Logistics
CPI	Continuous process improvement
CPM	Continuous Parsing Measure
CPN	Colored Petri net
CPSI	Continual process and service improvement
CRISP-DM	Cross Industry Standard Process for Data Mining
CRM	Customer relationship management
CSV	Comma-separated values
D2	Deliver make-to-order
DBMS	Database management system
EDI	Electronic Data Interchange
EEPC	Extended event-driven process chain
EFQM	European Foundation for Quality Management
EPC	Electronic product code
EPC	Event-driven process chain
EPCIS	Electronic product code information services
ERM	Entity relationship model
ERP	Enterprise Resource Planning
ETOM	Business Process Framework
Fitness	Token-Based Fitness
Fsc	Firing sequence compliance
Fscd	Firing sequence compliance degree
Fscm	Firing sequence compliance maturity
GCD	Greatest common divisor
GLN	Global Location Number
GRAI	Global Returnable Asset Identifier
GRC	Governance, risk, and compliance
GS1	Global Standard One
GUID	Globally Unique Identifier
HCL	Healthcare Logistics
HW	High Wycombe

IANA	Internet Assigned Numbers Authority
IAS	International Accounting Standard
IC	Interaction Center
ID	Identification
IDoc	Intermediate document
IEC	International Electrotechnical Commission
IS	Information system
ISO	International Organization for Standardization
ISP	Internet service provider
IT	Information technology
ITGI	IT Governance Institute
ITS CMM	IT Service Capability Maturity Model
ITSM	Information technology service management
Kent	Kent Pharmaceuticals
KGI	Key goal indicator
KonTraG	Gesetz zur Kontrolle und Transparenz im Unternehmensbereich
KPI	Key performance indicator
LCM	Least common multiple
Lcs	Longest common subsequence
LSP	Logistic service provider
LTL	Linear temporal logic
M2	Make-to-order
M_O_R	Management of Risk
MIS	Management information system
MRP	Material requirements planning
MXML	Mining Extensible Markup Language
NAR	Non-adherence to references
OEM	Original Equipment Manufacturer
OGC	Office of Government Commerce
OMG	Object Management Group
ONS	Object Name Service
PAIS	Process-aware information system
PC	Personal computer

Acronyms

PCF	APQC Process Classification Framework SM
PLM	Product life cycle management
PML	Product Markup Language
PNML	Petri Net Markup Language
ProM	Process mining framework
RFC	Remote function call
RfC	Request for change
RFID	Radio frequency identification
RTI	Returnable item
S2	Source Make-to-Order
SAM	Strategic Alignment Model
SCC	Supply Chain Operations Council
SCM	Supply Chain Management
SCOR	Supply Chain Operations Reference-Model
SD	Standard deviation
SGTIN	Serialized Global Trade Item Number
SOA	Service-Oriented Architecture
SOX	Sarbanes-Oxley Act
SRM	Supplier Relationship Management
SSCC	Serial Shipping Container Code
StrApp	Advanced structural appropriateness
TOGAF	The Open Group Architecture Framework
TQM	Total quality management
TSO	The Stationery Office
URI	Uniform Resource Identifier
URN	Uniform Resource Name
WF-net	Workflow net
WFMS	Workflow management system
WIP	Work-in-progress
XML	Extensible Markup Language

Acknowledgment

I could never have completed this dissertation without the support and encouragement of many persons.

In particular, I wish to thank and express my deep and sincere gratitude to my supervisor Professor Oliver Günther for his guidance over the years. He inspired and directed me with advice and insightful comments throughout my doctoral research and writing and provided a superb research environment at the Institute of Information Systems. Thanks as well to my second advisor, Professor Myra Spiliopoulou, for her support. Collaborating with her and her staff has been a pleasure for me. My colleagues at the Institute of Information Systems, including Professor Jan Mendling and Professor Gerrit Tamm, zeally cooperated. Professor Mendling and I carried out empirical research into process mining of RFID-based supply chains. Professor Tamm and I undertook extensive research into the continual improvement of processes and services, and he facilitated many corporate contacts. Indeed, Professor Tamm provided inspiration throughout the life of this dissertation.

This work here has been supported and funded by SAP Research Dresden. I am grateful for the personal freedom of working at home granted by Dr. Uwe Kubach. Dr. Gregor Hackenbroich aided me in preparing for the examinations prerequisite to admission to the doctoral program. Professor Jorge Cardoso guided me in researching compliance with reference models; his encouragement, advice, and criticism were indispensable. I would also like to recognize all of my fellow doctoral students and the administrators with whom I have had the pleasure of working during my time at SAP. To my great delight, I supervised and worked with Alexander Claus, who did brilliant graduate work around the topic of my dissertation and who assiduously strove to turn the concept into a working prototype.

Konstantin Petruch offered me valuable information and support regarding the continual process improvement approach – our discussions have always been very helpful.

James J. Toomey gave me invaluable editing suggestions that shaped the manuscript.

I want to thank my family for its invariable support, love, care, and encouragement. My parents provided me with unconditional love and support throughout my long education; they truly understood and encouraged me in all my decisions. My father sacrificed an enormous amount of time proof-reading this dissertation. My husband parents' assistance has been absolutely essential.

My sons Marius and Philipp never left my mind as I pursued this goal. I hope and trust that my attainments will be matched and then exceeded by them. They were prodigiously patient with their mom.

Acknowledgment

I owe profound thanks to my loving husband Marcus. Not only did he offer invaluable advice and encouragement – he took care of *everything*. I am blessed to have him as my life partner. He has sustained me and this work. I dedicate this dissertation to him.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Fröndenberg, den 07.02.2011

Kerstin Gerke

